

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:35:42 ; Search time 44.9 Seconds
(without alignments)
950.215 Million cell updates/sec

Title: US-09-543-407-12

Perfect score: 779
Sequence: 1 MKLKVAAFAAIVVSGSALA.....VTRVTHEMAHANNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04.*

- 1: Geneseq1980s.*
- 2: Geneseq1990s.*
- 3: Geneseq2000s.*
- 4: Geneseq2001s.*
- 5: Geneseq2002s.*
- 6: Geneseq2003as.*
- 7: Geneseq2003bs.*
- 8: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	779	100.0	151	3 AAB36346	Aab36346 Agfa::PT3
2	696	89.3	151	3 AAB36347	Aab36347 Agfa::PT3
3	695	89.2	151	2 AAR74625	Aar74625 Agfa sequ
4	695	89.2	151	3 AAB36341	Aab36341 Salmonell
5	690	88.6	151	2 AAW23570	Aaw23570 Salmonell
6	655	84.1	151	3 AAB36352	Aab36352 Agfa::PT3
7	614	78.8	151	3 AAB36353	Aab36353 Agfa::PT3
8	613	78.7	151	3 AAB36349	Aab36349 Agfa::PT3
9	611	78.4	151	3 AAB36350	Aab36350 Agfa::PT3
10	605	77.7	151	3 AAB36354	Aab36354 Agfa::PT3
11	604	77.5	151	3 AAB36351	Aab36351 Agfa::PT3
12	603	77.4	151	3 AAB36355	Aab36355 Agfa::PT3
13	580	74.5	151	3 AAB36348	Aab36348 Agfa::PT3
14	560	71.9	120	2 AAR62761	Aar62761 Agfa sequ
15	560	71.9	120	2 AAW23569	Aaw23569 Salmonell
16	520	66.8	151	3 AAB36343	Aab36343 Escherich
17	515	66.1	151	7 ABR82651	AbR82651 E. coli C
18	485	62.3	142	2 AAR52664	Aar52664 Fibronect
19	413	53.0	122	2 AAR52663	Aar52663 FNB curli
20	237	30.4	45	3 AAB36316	Aab36316 Salmonell
21	132	16.9	22	3 AAB36318	Aab36318 Salmonell
22	123	15.8	23	3 AAB36321	Aab36321 Salmonell
23	123	15.8	23	3 AAB36326	Aab36326 Salmonell
24	123	15.8	23	3 AAB36338	Aab36338 Salmonell
25	115	14.8	22	3 AAB36325	Aab36325 Salmonell

26	115	14.8	22	3 AAB36339	Aab36339 Salmonell
27	115	14.8	22	3 AAB36320	Aab36320 Salmonell
28	113	14.5	24	7 ABR82644	AbR82644 E. coli C
29	109	14.0	23	3 AAB36340	Aab36340 Salmonell
30	109	14.0	23	3 AAB36324	Aab36324 Salmonell
31	109	14.0	23	3 AAB36319	Aab36319 Salmonell
32	100.5	12.9	151	3 AAB36342	Aab36342 Salmonell
33	98.5	12.6	850	4 ABB65764	Abb65764 Drosophil
34	98.5	12.6	1028	4 ABB62708	Abb62708 Drosophil
35	98	12.6	26	7 ABR82645	AbR82645 E. coli C
36	95	12.2	24	7 ABR82647	AbR82647 E. coli C
37	94.5	12.1	678	6 ABU36649	AbU36649 Protein e
38	94.5	12.1	688	5 ABP74039	AbP74039 Candida a
39	93	11.9	764	6 AAE36890	Aae36890 Plectreir
40	92	11.8	23	3 AAB36331	Aab36331 Escherich
41	92	11.8	673	3 AAY44403	Aay44403 Human tru
42	92	11.8	673	5 AAU79538	Aau79538 Truncated
43	92	11.8	949	3 AAY44404	Aay44404 Human tru
44	92	11.8	949	5 AAU79539	Aau79539 Truncated
45	92	11.8	1327	3 AAY44402	Aay44402 Human tan

ALIGNMENTS

RESULT 1

AAB36346
ID AAB36346 standard; protein; 151 AA.
XX
AC AAB36346;
XX
XX 26-FEB-2001 (first entry)
XX
XX Agfa::PT3#1 amino acid sequence SEQ ID NO:12.
XX
XX Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
XX vaccine; immune response; immunogen.
XX
XX Salmonella enteritidis.
XX
XX Escherichia coli.
XX Synthetic.
XX
XX WO2000060102-A2.
XX
XX 12-OCT-2000.
XX
XX 05-APR-2000; 2000WO-CA000356.
XX
XX 05-APR-1999; 99US-0127888P.
XX
XX (UYVI-) UNIV VICTORIA.
XX
XX White AP, Doran JL, Collison SK, Kay MW;
XX WPI; 2000-672631/65.
XX N-PSDB; AAC64622.
XX

Recombinant agfa gene having a segment replaced by a foreign DNA sequence which encodes foreign epitope or antigen, expresses recombinant Agfa protein useful for eliciting immune response in animal.

Disclosure; Page 135; 139pp; English.

The present invention describes a recombinant agfa gene (1) where a segment of the gene has been replaced by a segment of a foreign DNA sequence which encodes a foreign epitope or antigen. Also described are: (1) use of thin aggregative fimbriae (SEPI7/TAF) nucleation depended assembly system of strains of Salmonella, Escherichia coli and Enterobacteriaceae for the production of fimbriae comprising recombinant Agfa, CsgA, and Agfa-homologue fimbrin subunits, respectively; (2) directing recombination of a recombinant gene into the chromosome of the homologous species; (3) directing recombination of a recombinant gene back into the chromosome of the homologous species, replacing the native

CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant AgfA
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX Sequence 151 AA;
 SQ

Query Match 100.0%; Score 779; DB 3; Length 151;
 Best Local Similarity 100.0%; Pred. No. 2e-66; 0; Indels 0; Gaps 0;
 Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
 DB 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
 QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGG 120
 DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGG 120
 QY 121 NNAALVNYDQLVTRVVTHEMAHANNTATQY 151
 DB 121 NNAALVNYDQLVTRVVTHEMAHANNTATQY 151

RESULT 2
 AAB36347
 ID AAB36347 standard; protein; 151 AA.
 AC AAB36347;
 XX
 XX 26-FEB-2001 (first entry)
 DT
 DE AgfA::PT3#2 amino acid sequence SEQ ID NO:14.
 DE
 KW Salmonella; agfA; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 XX WO2000060102-A2.
 PN
 PD 12-OCT-2000.
 XX
 XX 05-APR-2000; 2000WO-CA000356.
 PF
 XX
 XX 05-APR-1999; 99US-0127888P.
 PR
 XX
 XX (UYVI-) UNIV VICTORIA.
 PA
 XX White AP, Doran JL, Collison SK, Kay WW;
 PI
 XX WPI; 2000-672631/65.
 DR
 DR N-PSDB; AAC64623.
 XX
 XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant AgfA
 PT protein useful for eliciting immune response in animal.
 XX
 XX Disclosure; Page 136; 139pp; English.
 PS

XX The present invention describes a recombinant agfA gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAf) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant AgfA
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX Sequence 151 AA;
 SQ

Query Match 89.3%; Score 696; DB 3; Length 151;
 Best Local Similarity 87.6%; Pred. No. 1.7e-58;
 Matches 141; Conservative 0; Mismatches 0; Indels 20; Gaps 2;
 QY 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
 DB 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
 QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGG 120
 DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGG 120
 QY 121 NNAALVNYDQLVTRVVTHEMAHANNTATQY 151
 DB 118 -----YDQLVTRVVTHEMAHANNTATQY 151

RESULT 3
 AAR74625
 ID AAR74625 standard; protein; 151 AA.
 XX
 XX AAR74625;
 XX
 XX 25-MAR-2003 (revised)
 DT 26-JUN-1995 (first entry)
 DT
 XX AgfA sequence.
 DE
 XX Salmonella; AgfA; vaccine.
 KW
 XX Salmonella.
 OS
 XX WO9425598-A2.
 PN
 PD 10-NOV-1994.
 XX
 XX 26-APR-1994; 94WO-IB000207.
 PF
 XX
 XX 26-APR-1993; 93US-00054452.
 PR
 XX (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
 PA (KING/) KING J.
 XX

PI Kay WW, Collinson SK, Clouthier SC, Doran JL;
XX WPI; 1994-358275/44.
DR N-PSDB; AAQ87467.
XX
XX Eliciting an immune response to Salmonella - using attenuated Salmonella
PT strains, vector constructs, or compenss. contg. fimbrial type proteins.
XX
XX Disclosure; Fig 7B; 95pp; English.
XX
XX The Salmonella Agfa protein and DNA are used in vaccine and genetic
CC immunization compositions, respectively, to elicit an immune response to
CC Salmonella in animals (e.g. food producing animals) and humans. (Updated
CC on 25-MAR-2003 to correct PN field.)
XX
SQ Sequence 151 AA;
Query Match 89.2%; Score 695; DB 2; Length 151;
Best Local Similarity 90.1%; Pred. No. 2.2e-58;
Matches 136; Conservative 4; Mismatches 11; Indels 0; Gaps 0;
QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
QY 121 NNAALVNDQLVTRVVTTHMAHANANATANY 151
DB 121 NNAALVNQTASDSSVMVRQVGFNNATANY 151
RESULT 4
AAB36341
ID AAB36341 standard; protein; 151 AA.
XX
XX AAB36341;
XX
DT 26-FEB-2001 (first entry)
XX
DE Salmonella enteritidis Agfa amino acid sequence SEQ ID NO:5.
XX
XX Salmonella; agfa; chromosomal gene replacement; fimbrian; epitope;
KW vaccine; immune response; immunogen.
XX
OS Salmonella enteritidis.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
XX 05-APR-2000; 2000WO-CA000356.
PF
XX 05-APR-1999; 99US-0127888P.
PR
XX (UYVI-) UNIV VICTORIA.
PA
XX White AP, Doran JL, Collinson SK, Kay WW;
PI
XX WPI; 2000-672631/55.
DR
DR N-PSDB; AAC64617.
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 135; 139pp; English.
XX
XX The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:

CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella. Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbrian subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrian protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 151 AA;
Query Match 89.2%; Score 695; DB 3; Length 151;
Best Local Similarity 90.1%; Pred. No. 2.2e-58;
Matches 136; Conservative 4; Mismatches 11; Indels 0; Gaps 0;
QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
QY 121 NNAALVNDQLVTRVVTTHMAHANANATANY 151
DB 121 NNAALVNQTASDSSVMVRQVGFNNATANY 151
RESULT 5
AAW23570
ID AAW23570 standard; protein; 151 AA.
XX
XX AAW23570;
XX
DT 25-MAR-2003 (revised)
DT 29-SEP-1997 (first entry)
XX
DE Salmonella enteritidis 27655-3b agfa.
XX
KW Enteropathogenic bacteria; enterobacteria; S. enteritidis; antibody.
XX
OS Salmonella enteritidis.
XX
XX Key Location/Qualifiers
FH Misc-difference 123
FT /note= "Encoded by GCC"
XX
XX US5635617-A.
XX
XX 03-JUN-1997.
XX
XX 26-APR-1994; 94US-00233788.
XX
XX 26-APR-1993; 93US-00054452.
XX
XX (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX
XX Collinson SK, Kay WW, Doran JL;

```

XX WPI: 1997-309886/28.
XX N-PSDB; AAT74142.
XX
XX Isolated Salmonella gene agfa - used for diagnosis of Salmonella or
XX enteropathogenic bacteria of the Enterobacteriaceae family.
XX
XX Example 2; Fig 7; 85pp; English.
XX
XX The present sequence represents agfa encoded by the full agfa gene
XX derived from Salmonella enteritidis 27655-3b. The nucleic acid can be
XX used to provide diagnostic assays for Salmonella and/or enteropathogenic
XX bacteria of the family Enterobacteriaceae. It can also be used to provide
XX proteins and antibodies which can be used for assays. The nucleic acid
XX sequence can be used to provide probes or primers which can specifically
XX hybridise to nucleic acid molecules from greater than 99% of Salmonella
XX strains that are pathogenic to warm-blooded animals relative to nucleic
XX acid molecules from virtually all other microbial organisms. (Updated on
XX 25-MAR-2003 to correct PF field.)
XX
XX Sequence 151 AA;
SQ
Query Match 88.6%; Score 690; DB 2; Length 151;
Best Local Similarity 89.4%; Pred. No. 6.5e-58;
Matches 135; Conservative 4; Mismatches 12; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGOYGG 120
DB 61 SDARKSETTITQSGYNGADVGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGOYGG 120
QY 121 NNAALVNYDQVTRVVTHEMAHANQATSDSSVMVRQVGFNNATANQY 151
DB 121 NNPALVNYQTASDSSVMVRQVGFNNATANQY 151
RESULT 6
ID AAB36352 standard; protein; 151 AA.
XX
AC AAB36352;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa::PT3#7 amino acid sequence SEQ ID NO:24.
XX
KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WC200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
XX WPI: 2000-672631/65.
XX N-PSDB; AAC64628.
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
XX which encodes foreign epitope or antigen, expresses recombinant Agfa

```

```

PT protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 138; 139pp; English.
XX
XX The present invention describes a recombinant agfa gene (I) where a
XX segment of the gene has been replaced by a segment of a foreign DNA
XX sequence which encodes a foreign epitope or antigen. Also described are:
XX (1) use of thin aggregative fimbriae (SEF17/7AF) nucleation depended
XX assembly system of strains of Salmonella, Escherichia coli and
XX Enterobacteriaceae for the production of fimbriae comprising recombinant
XX Agfa, CsgA and Agfa-homologue fimbrin subunits, respectively; (2)
XX directing recombination of a recombinant gene into the chromosome of the
XX homologous species; (3) directing recombination of a recombinant gene
XX back into the chromosome of the homologous species, replacing the native
XX copy of that gene; and (4) eliciting an immune response in an animal,
XX comprising separating an amino acid polymer comprising a recombinant Agfa
XX protein containing a replacement segment or segments of foreign amino
XX acid sequence or sequences grown on a Salmonella, E. coli or
XX Enterobacteriaceae host cell, from the host cell and introducing the
XX polymer into the animal in conjunction with a carrier or diluent. (I) is
XX useful for the expression of recombinant Agfa protein which is useful for
XX eliciting an immune response in an animal. In a fimbrial presentation
XX system the heterologous antigens are presented in high numbers (up to
XX 500,000 copies/cell), the hybrid fimbrin protein possesses both the
XX immunogenicity and adhesion properties relevant for an efficient live
XX vaccine, the carrier fimbrial subunit proteins are usually strong
XX immunogens, which may be important for directing an immune response
XX against the inserted epitope, and hybrid fimbriae are easy and
XX inexpensive to purify in large amount. The present sequence is given in
XX the exemplification of the present invention
XX
XX Sequence 151 AA;
Query Match 84.1%; Score 655; DB 3; Length 151;
Best Local Similarity 79.8%; Pred. No. 1.4e-54;
Matches 134; Conservative 0; Mismatches 0; Indels 34; Gaps 2;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGOYGG 120
DB 61 SDARKSETTITQSGYNGADVGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGOYGG 120
QY 121 NNAALVNYDQVTRVVTHEMAHA-----NNATANQY 151
DB 110 -----NYDQVTRVVTHEMAHANQATSDSSVMVRQVGFNNATANQY 151
RESULT 7
AAB36353
ID AAB36353 standard; protein; 151 AA.
XX
AC AAB36353;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa::PT3#8 amino acid sequence SEQ ID NO:26.
XX
KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WC200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
XX

```

```
PR 05-APR-1999; 99US-0127888P.
XX (UYVI-) UNIV VICTORIA.
PA White AP, Doran JL, Collison SK, Kay WW;
XX WPI; 2000-672631/65.
XX N-PSDB; AAC64629.
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX Disclosure; Page 136; 139pp; English.
XX The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAP) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX Sequence 151 AA;
SQ
Query Match 78.8%; Score 614; DB 3; Length 151;
Best Local Similarity 80.8%; Pred. No. 1.2e-50;
Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALYDQ 60
QY 61 SDARKSETTITQSGYNGADYVGGADNSTIETQNGFRNATIDQWNAKNSDITVGYGG 120
DB 61 LVTRVVTHEMAHAGYNGADYVGGADNSTIETQNGFRNATIDQWNAKNSDITVGYGG 120
QY 121 NNAALVNYDQLVTRVVTHEMAHANNATANCY 151
DB 121 NNAALVNYDQLVTRVVTHEMAHANNATANCY 151
RESULT 8
ID AAB36349 standard; protein; 151 AA.
XX AAB36349;
AC AAB36349;
XX 26-FEB-2001 (first entry)
XX Agfa::PT3#4 amino acid sequence SEQ ID NO:18.
XX Salmonella; agfa; chromosomal gene replacement; fimbrial; epitope;
KW vaccine; immune response; immunogen.
```

```
XX Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX WO2000060102-A2.
XX 12-OCT-2000.
XX 05-APR-2000; 2000WO-CA000356.
XX 05-APR-1999; 99US-0127888P.
XX (UYVI-) UNIV VICTORIA.
XX White AP, Doran JL, Collison SK, Kay WW;
XX WPI; 2000-672631/65.
XX N-PSDB; AAC64625.
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX Disclosure; Page 136; 139pp; English.
XX The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAP) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX Sequence 151 AA;
SQ
Query Match 78.7%; Score 613; DB 3; Length 151;
Best Local Similarity 80.8%; Pred. No. 1.5e-50;
Matches 122; Conservative 7; Mismatches 22; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDYDQLVTRVVTHEMAHALQ 60
QY 61 SDARKSETTITQSGYNGADYVGGADNSTIETQNGFRNATIDQWNAKNSDITVGYGG 120
DB 61 SDARKSETTITQSGYNGADYVGGADNSTIETQNGFRNATIDQWNAKNSDITVGYGG 120
QY 121 NNAALVNYDQLVTRVVTHEMAHANNATANCY 151
DB 121 NNAALVNYDQLVTRVVTHEMAHANNATANCY 151
RESULT 9
```


CC the exemplification of the present invention
 XX Sequence 151 AA;
 SQ

Query Match 77.7%; Score 605; DB 3; Length 151;
 Best Local Similarity 80.1%; Pred. No. 8.5e-50;
 Matches 121; Conservative 5; Mismatches 25; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALQ 60
 QY 61 SDARKSETTITQSGYNGADYQGGADNSTIELTQNGFRNNATIDQWNAKNSDIITVGYGG 120
 DB 61 SDARKSETTITQSGYNGADYQGGADNSTIELTQNGFRNNATIDQWNAKNSDIITVGYGG 120
 QY 121 NNAALVNYDQVLRVVTTHMAHANNATANQY 151
 DB 121 NNAALVNYDQVLRVVTTHMAHANNATANQY 151

RESULT 11
 AAB36351
 ID AAB36351 standard; protein; 151 AA.
 XX
 AC AAB36351;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Agfa::PT3#6 amino acid sequence SEQ ID NO:22.
 XX
 KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 PN WO2000060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay WW;
 XX
 DR WPI: 2000-672631/65.
 DR N-PSDB; AAC64627.
 XX
 CC Recombinant agfa gene having a segment replaced by a foreign DNA sequence which encodes foreign epitope or antigen, expresses recombinant Agfa protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 137; 139pp; English.
 XX

The present invention describes a recombinant agfa gene (I) where a segment of the gene has been replaced by a segment of a foreign DNA sequence which encodes a foreign epitope or antigen. Also described are:
 (1) use of thin aggregative fimbriae (SPF17/TAP) nucleation depended assembly system of strains of Salmonella, Escherichia coli and Enterobacteriaceae for the production of fimbriae comprising recombinant Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2) directing recombination of a recombinant gene into the chromosome of the homologous species; (3) directing recombination of a recombinant gene back into the chromosome of the homologous species, replacing the native copy of that gene; and (4) eliciting an immune response in an animal comprising separating an amino acid polymer comprising a recombinant Agfa protein containing a replacement segment or segments of foreign amino acid sequence or sequences grown on a Salmonella, E. coli or

CC Enterobacteriaceae host cell, from the host cell and introducing the polymer into the animal in conjunction with a carrier or diluent. (I) is useful for the expression of recombinant Agfa protein which is useful for eliciting an immune response in an animal. In a fimbrial presentation system the heterologous antigens are presented in high numbers (up to 500,000 copies/cell), the hybrid fimbrin protein possesses both the immunogenicity and adhesion properties relevant for an efficient live vaccine, the carrier fimbrial subunit proteins are usually strong immunogens, which may be important for directing an immune response against the inserted epitope, and hybrid fimbriae are easy and inexpensive to purify in large amount. The present sequence is given in the exemplification of the present invention
 XX Sequence 151 AA;
 SQ

Query Match 77.5%; Score 604; DB 3; Length 151;
 Best Local Similarity 80.8%; Pred. No. 1.1e-49;
 Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALQ 60
 QY 61 SDARKSETTITQSGYNGADYQGGADNSTIELTQNGFRNNATIDQWNAKNSDIITVGYGG 120
 DB 61 SDARKSETTITQSGYNGADYQGGADNSTIELTQNGFRNNATIDQWNAKNSDIITVGYGG 120
 QY 121 NNAALVNYDQVLRVVTTHMAHANNATANQY 151
 DB 121 NNAALVNYDQVLRVVTTHMAHANNATANQY 151

RESULT 12
 AAB36355
 ID AAB36355 standard; protein; 151 AA.
 XX
 AC AAB36355;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Agfa::PT3#10 amino acid sequence SEQ ID NO:30.
 XX
 KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 PN WO2000060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay WW;
 XX
 DR WPI: 2000-672631/65.
 DR N-PSDB; AAC64631.
 XX
 CC Recombinant agfa gene having a segment replaced by a foreign DNA sequence which encodes foreign epitope or antigen, expresses recombinant Agfa protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 139; 139pp; English.
 XX

The present invention describes a recombinant agfa gene (I) where a segment of the gene has been replaced by a segment of a foreign DNA sequence which encodes a foreign epitope or antigen. Also described are:
 (1) use of thin aggregative fimbriae (SPF17/TAP) nucleation depended assembly system of strains of Salmonella, Escherichia coli and Enterobacteriaceae for the production of fimbriae comprising recombinant Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2) directing recombination of a recombinant gene into the chromosome of the homologous species; (3) directing recombination of a recombinant gene back into the chromosome of the homologous species, replacing the native copy of that gene; and (4) eliciting an immune response in an animal comprising separating an amino acid polymer comprising a recombinant Agfa protein containing a replacement segment or segments of foreign amino acid sequence or sequences grown on a Salmonella, E. coli or

CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC AgfA, CsgA, and AgfA-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant AgfA
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX Sequence 151 AA;
 SQ

Query Match 77.4%; Score 603; DB 3; Length 151;
 Best Local Similarity 80.1%; Pred. No. 1.3e-49;
 Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;
 QY 1 MKLLKVAAPAAIIVSGSALAGVVPQGGGNGHNGGNSGGPSTLSIYQYGSANAALALQ 60
 DB 1 MKLLKVAAPAAIIVSGSALAGVVPQGGGNGHNGGNSGGPSTLSIYQYGSANAALALQ 60
 QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNATIDQWNAKNSDITVQYGG 120
 DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNATIDQWNAKNSDITVQYGG 120
 QY 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151
 DB 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151

RESULT 13
 AAB36348
 ID AAB36348 standard; protein; 151 AA.
 XX
 AC AAB36348;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE AgfA::P3#3 amino acid sequence SEQ ID NO:16.
 XX
 KW Salmonella; agfA; chromosomal gene replacement; fimbrial; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 PN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UTVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay MW;
 XX WPI; 2000-672631/65.
 XX
 PD

DR N-PSDB; AAC64624.
 XX
 PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant AgfA
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 136; 139pp; English.

XX The present invention describes a recombinant agfA gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC AgfA, CsgA, and AgfA-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant AgfA
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX Sequence 151 AA;

Query Match 74.5%; Score 580; DB 3; Length 151;
 Best Local Similarity 80.1%; Pred. No. 2.1e-47;
 Matches 121; Conservative 7; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIIVSGSALAGVVPQGGGNGHNGGNSGGPSTLSIYQYGSANAALALQ 60
 DB 1 MKLLKVAAPAAIIVSGSALAGVVPQGGGNGHNGGNSGGPSTLSIYQYGSANAALALQ 60
 QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNATIDQWNAKNSDITVQYGG 120
 DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNATIDQWNAKNSDITVQYGG 120
 QY 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151
 DB 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151

RESULT 14
 AAR62761
 ID AAR62761 standard; protein; 120 AA.
 XX
 AC AAR62761;
 XX
 DT 27-AUG-2003 (revised)
 DT 25-MAR-2003 (revised)
 DT 26-JUN-1995 (first entry)
 XX
 DE AgfA sequence.
 XX
 KW Salmonella; AgfA; vaccine.
 XX
 OS Salmonella enteritidis.
 XX
 PN WO9425598-A2.
 XX
 PD 10-NOV-1994.

```

XX 26-APR-1994; 94WO-IB000207.
XX
XX 26-APR-1993; 93US-00054452.
XX
XX (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX (KING/) KING J.
XX
XX Kay WW, Collinson SK, Clouthier SC, Doran JL;
XX
XX WPI; 1994-358275/44.
XX N-PSDB; AAQ73066.
XX
XX Eliciting an immune response to Salmonella - using attenuated Salmonella
XX strains, vector constructs, or compens. contg. fimbrial type proteins.
XX
XX Disclosure; Fig 7A; 95pp; English.
XX
XX The sequence represents the Salmonella enteritis 27655-3b TnpHoA mutant
XX strain Agfa protein. The encoding DNA and isolated Agfa protein are used
XX in genetic immunization and vaccine compositions, respectively, to elicit
XX an immune response to Salmonella in animals (e.g. food producing animals)
XX and humans. (Updated on 25-MAR-2003 to correct PN field.) (Updated on 27-
XX AUG-2003 to correct OS field.)
XX
XX Sequence 120 AA;
SQ
Query Match 71.9%; Score 560; DB 2; Length 120;
Best Local Similarity 100.0%; Pred. No. 1.2e-45;
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 22 VVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGADV 81
DB 1 VVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGADV 60
QY 82 GQADNSTIETQNGFRNNATIDQWNAKNSDITVGYGGNNAALVN 127
DB 61 GQADNSTIETQNGFRNNATIDQWNAKNSDITVGYGGNNAALVN 106
RESULT 15
AAW23569
ID AAW23569 standard; protein; 120 AA.
XX
XX AAW23569;
XX
XX 25-MAR-2003 (revised)
XX 29-SEP-1997 (first entry)
XX
XX Salmonella enteritis 27655-3b TnpHoA mutant agfa fragment.
XX Enteropathogenic bacteria; enterobacteria; S. enteritis; antibody.
XX Salmonella enteritis.
XX
XX US5635617-A.
XX
XX 03-JUN-1997.
XX
XX 26-APR-1994; 94US-00233788.
XX
XX 26-APR-1993; 93US-00054452.
XX
XX (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX Collinson SK, Kay WW, Doran JL;
XX
XX WPI; 1997-309886/28.
XX N-PSDB; AAT74141.
XX
XX Isolated Salmonella gene agfa - used for diagnosis of Salmonella or
XX enteropathogenic bacteria of the Enterobacteria family.
XX

```

```

PS Example 2; Fig 7; 85pp; English.
XX
XX The present sequence represents an agfa fragment encoded by an agfa gene
XX fragment derived from Salmonella enteritis 27655-3b TnpHoA mutant
XX strain. The nucleic acid can be used to provide diagnostic assays for
XX Salmonella and/or enteropathogenic bacteria of the family Enterobacteria.
XX It can also be used to provide proteins and antibodies which can be used
XX for assays. The nucleic acid sequence can be used to provide probes or
XX primers which can specifically hybridize to nucleic acid molecules from
XX greater than 99% of Salmonella strains that are pathogenic to warm-
XX blooded animals relative to nucleic acid molecules from virtually all
XX other microbial organisms. (Updated on 25-MAR-2003 to correct PF field.)
XX
XX Sequence 120 AA;
SQ
Query Match 71.9%; Score 560; DB 2; Length 120;
Best Local Similarity 100.0%; Pred. No. 1.2e-45;
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 22 VVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGADV 81
DB 1 VVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGADV 60
QY 82 GQADNSTIETQNGFRNNATIDQWNAKNSDITVGYGGNNAALVN 127
DB 61 GQADNSTIETQNGFRNNATIDQWNAKNSDITVGYGGNNAALVN 106
Search completed: August 2, 2004, 14:48:24
Job time : 45.9 secs

```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:40:48 ; Search time 12 Seconds
(without alignments)
649.627 Million cell updates/sec

Title: US-09-543-407-12

Perfect score: 779
Sequence: 1 MKLLKVAFAAIVVSGSALA.....VTRVVTHEMAHANNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued Patents AA.*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	690	88.6	151	1	US-08-233-788A-59
2	560	71.9	120	1	US-08-233-788A-57
3	92	11.8	673	3	US-09-196-387-8
4	92	11.8	673	4	US-09-841-835-8
5	92	11.8	949	3	US-08-196-387-10
6	92	11.8	949	4	US-09-841-835-10
7	92	11.8	1327	3	US-09-196-387-2
8	92	11.8	1327	4	US-09-841-835-2
9	92	11.8	1327	4	US-09-972-115A-8
10	90.5	11.6	745	4	US-09-336-115C-6
11	89.5	11.5	738	3	US-08-864-038A-3
12	89	11.4	943	4	US-09-056-556-204
13	89	11.4	943	4	US-09-072-596-199
14	89	11.4	943	4	US-09-477-135A-131
15	89	11.4	943	4	US-09-072-967-204
16	87.5	11.2	892	4	US-09-336-447A-5
17	85	10.9	956	4	US-09-134-078-63
18	82.5	10.6	186	4	US-09-382-276-3
19	82.5	10.6	873	4	US-09-336-447A-13
20	81.5	10.5	339	4	US-09-252-991A-32096
21	81	10.4	568	4	US-09-543-681A-6965
22	81	10.4	1739	4	US-09-540-236-3739
23	81	10.4	1864	2	US-08-804-227C-3
24	80	10.3	232	4	US-09-252-921A-30263
25	80	10.3	975	4	US-09-328-352-4764
26	80	10.3	3241	4	US-09-841-786-1
27	79.5	10.2	361	4	US-09-540-236-2164

28	79	10.1	941	4	US-09-336-447A-9	Sequence 9, Appli
29	78.5	10.1	2123	3	US-08-968-685A-10	Sequence 10, Appli
30	78	10.0	306	1	US-08-254-573-2	Sequence 2, Appli
31	78	10.0	306	1	US-08-687-379-4	Sequence 2, Appli
32	78	10.0	306	1	US-08-687-379-4	Sequence 4, Appli
33	78	10.0	306	4	US-08-172-332-1	Sequence 1, Appli
34	78	10.0	306	4	US-08-216-326-2	Sequence 2, Appli
35	77.5	9.9	702	4	US-09-252-991A-22119	Sequence 22119, A
36	77.5	9.9	714	4	US-09-841-786-4	Sequence 4, Appli
37	77.5	9.9	878	4	US-09-540-236-3401	Sequence 3401, Ap
38	77	9.9	415	4	US-09-025-769B-280	Sequence 280, App
39	77	9.9	528	4	US-09-490-291-8	Sequence 8, Appli
40	77	9.9	1690	4	US-09-595-684B-39	Sequence 39, Appli
41	77	9.9	2315	4	US-09-543-681A-5434	Sequence 5434, Ap
42	76.5	9.8	159	3	US-08-856-253-2	Sequence 2, Appli
43	76.5	9.8	186	4	US-09-382-276-2	Sequence 2, Appli
44	76.5	9.8	605	4	US-09-489-039A-13002	Sequence 13002, A
45	75.5	9.7	624	4	US-09-336-447A-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1
US-08-233-788A-59
; Sequence 59, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:

APPLICANT: Doran, James L.
APPLICANT: Key, William W.
APPLICANT: Collinson, Karen S.
APPLICANT: Clouthier, Sharon C.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
TITLE OF INVENTION: OF SALMONELLA
NUMBER OF SEQUENCES: 61
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed and Berry
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: U.S.A.
ZIP: 98104-7092

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICANT NUMBER: US/08/233,788A
FILING DATE: 26-APR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: King, Joshua
REGISTRATION NUMBER: 35,570
REFERENCE/DOCKET NUMBER: 920043.403C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEEDANBERRY
INFORMATION FOR SEQ ID NO: 59:
SEQUENCE CHARACTERISTICS:
LENGTH: 151 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: Protein
US-08-233-788A-59

Query Match 88.6%; Score 690; DB 1; Length 151;
Best Local Similarity 89.4%; Pred. No. 7, 1e-62;
Matches 135; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALGVVPQWGGNGNGSGPSTLSIYQYGSANAALQ 60
|||||

Db 1 MKLLKVAFAAIVVSGSALAGVVPWQGGGNGHNGGNSGPDSTLSIYQGSANAALQ 60
QY 61 SDARKSETTITQSGYNGADVGQGDNSTIETQNGFRNNATIDQWNAKNSDITVQYGG 120
Db 61 SDARKSETTITQSGYNGADVGQGDNSTIETQNGFRNNATIDQWNAKNSDITVQYGG 120
QY 121 NNAALVNYDQLVTRVVTTHMAHANNATANY 151
Db 121 NNPALVNOTASDSSVMVRQVGFNNATANY 151

RESULT 2
US-08-233-788A-57
; Sequence 57, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Key, William W.
; APPLICANT: Collinson, Karen S.
; APPLICANT: Clouthier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,788A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERY
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 120 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-233-788A-57

Query Match 71.9%; Score 560; DB 1; Length 120;
Best Local Similarity 100.0%; Pred. No. 6e-49;
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 22 VVPQGGGNGHNGGNSGPDSTLSIYQGSANAALQSDARKSETTITQSGYNGADV 81
Db 1 VVPQGGGNGHNGGNSGPDSTLSIYQGSANAALQSDARKSETTITQSGYNGADV 60
QY 82 GQGDNSTIETQNGFRNNATIDQWNAKNSDITVQYGGNNAALVN 127
Db 61 GQGDNSTIETQNGFRNNATIDQWNAKNSDITVQYGGNNAALVN 106

RESULT 3
US-09-196-387-8
; Sequence 8, Application US/09196387
; Patent No. 6277613
; GENERAL INFORMATION:

; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/196,387
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/095,225
; FILING DATE: June 10, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 673 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-196-387-8

Query Match 11.8%; Score 92; DB 3; Length 673;
Best Local Similarity 30.4%; Pred. No. 0.6;
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;
QY 6 VAAFAAI-VVSGSALAGVVPWQGGGNGHNGGNSGPDSTLSIYQGSANAALQSDAR 64
Db 99 VAAAPVPAVSTSSAAGVAPNPAGSGNNSPSSSSPTSS-SSSSPSPGSSLAESPAA 157
QY 65 KSETTIT----QSGYNGADVQGDNSTIETQNG--FRNNATIDQWNAKNSDI 113
Db 158 GVSSTAPLGPAGAGPGTGVPAVSGALRELLACRNGDVSRLVDAAVNNAKDM 212

RESULT 4
US-09-841-835-8
; Sequence 8, Application US/09841835
; Patent No. 6506587
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/841,835
/ FILING DATE:
/ CLASSIFICATION:
/ PRIOR APPLICATION DATA: 09/196,387
/ APPLICATION NUMBER: 09/196,387
/ FILING DATE:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Jackson Esq., David A.
/ REGISTRATION NUMBER: 26,742
/ REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 201-487-5800
/ TELEFAX: 201-343-1684
/ TELEX: 133521
/ INFORMATION FOR SEQ ID NO: 8:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 673 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-09-841-835-8

Query Match 11.8%; Score 92; DB 4; Length 673;
Best Local Similarity 30.4%; Pred. No. 0.6;
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

QY 6 VAAFAAI-VVGSALAGVFPQWGGGHHNGGSSGPDSTLSIVQGSANAIALQSDAR 64
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 99 VAAAPVVPVSTSSAAGVAPNPAGSGNNSSSSSPTSS-SSSSPSPGSSLAESPEAA 157
QY 65 KSTTTIT---QSGYNGADVGQCADNSTLTETLQNG--FRNATIDQWNAKNSDI 113
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 158 GVSTAPLPGCAAGGIGVPAVSGALRELLACRNGDVSRVKRLVDAAVNAKDM 212

RESULT 5
US-09-196-387-10
/ Sequence 10, Application US/09196387
/ Patent No. 6277613
/ GENERAL INFORMATION:
/ APPLICANT: de Lange, Titia
/ APPLICANT: Smith, Susan
/ TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
/ TITLE OF INVENTION: OF USE THEREOF
/ NUMBER OF SEQUENCES: 12
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Klauber & Jackson
/ STREET: 411 Hackensack Avenue, 4th Floor
/ CITY: Hackensack
/ STATE: New Jersey
/ COUNTRY: USA
/ ZIP: 07601
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/196,387
/ FILING DATE:
/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 09/095,225
/ FILING DATE: June 10, 1998
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Jackson Esq., David A.
/ REGISTRATION NUMBER: 26,742
/ REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 201-487-5800

```

Db 99 VAAAPVPAVSTSAAGVAPNPAGSGNNSPSSSSPTSS--SSSSPSPSGSLAESPEAA 157
Qy 65 KSETTIT---QSGYGNAGDVGGADNSTIETQNG--FRNNATIDQWNKNSDI 113
Db 158 GVSSTAPLPGGAAGPGTGPVAVSGALRELLEACRNGDVSRVKRLVDAANVNAKDM 212

RESULT 7

US-09-196-387-2
; Sequence 2, Application US/09196387
; Patent No. 6277613
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; TITLE OF INVENTION: OF USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/196,387
; FILING DATE:
; CLASSIFICATION:

PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/095,225
; FILING DATE: June 10, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521

INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1327 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO

Query Match 11.8%; Score 92; DB 3; Length 1327;
Best Local Similarity 30.4%; Pred. No. 1.5;
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

Qy 6 VAAFAAI-VVGSALAGVVPQWGGGNHNGNNSGPDSTLSIYQYGSANAALALQSDAR 64
Db 99 VAAAPVPAVSTSAAGVAPNPAGSGNNSPSSSSPTSS--SSSSPSPSGSLAESPEAA 157
Qy 65 KSETTIT---QSGYGNAGDVGGADNSTIETQNG--FRNNATIDQWNKNSDI 113
Db 158 GVSSTAPLPGGAAGPGTGPVAVSGALRELLEACRNGDVSRVKRLVDAANVNAKDM 212

RESULT 8

US-09-841-835-2
; Sequence 2, Application US/09841835
; Patent No. 6506597
; GENERAL INFORMATION:

; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; TITLE OF INVENTION: OF USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/841,835
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/196,387
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1327 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
US-09-841-835-2

Query Match 11.8%; Score 92; DB 4; Length 1327;
Best Local Similarity 30.4%; Pred. No. 1.5;
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

Qy 6 VAAFAAI-VVGSALAGVVPQWGGGNHNGNNSGPDSTLSIYQYGSANAALALQSDAR 64
Db 99 VAAAPVPAVSTSAAGVAPNPAGSGNNSPSSSSPTSS--SSSSPSPSGSLAESPEAA 157
Qy 65 KSETTIT---QSGYGNAGDVGGADNSTIETQNG--FRNNATIDQWNKNSDI 113
Db 158 GVSSTAPLPGGAAGPGTGPVAVSGALRELLEACRNGDVSRVKRLVDAANVNAKDM 212

RESULT 9

US-09-972-115A-8
; Sequence 8, Application US/09972115A
; Patent No. 6599728
; GENERAL INFORMATION:
; APPLICANT: Geron Corporation
; APPLICANT: Greg, Morin B.
; APPLICANT: Walter, Funk D.
; APPLICANT: Macyszewski, Piatyszek A.
; TITLE OF INVENTION: A Second Mammalian Telomerase
; FILE REFERENCE: 080/003C
; CURRENT APPLICATION NUMBER: US/09/972,115A
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: US 60/128,577
; PRIOR FILING DATE: 2000-04-10
; PRIOR APPLICATION NUMBER: US 60/129,123
; PRIOR FILING DATE: 1999-04-13
; NUMBER OF SEQ ID NOS: 64

```
; SOFTWARE: PatentIn version 3.11
; SEQ ID NO 8
; LENGTH: 1327
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-972-115A-8

Query Match      11.8%; Score 92; DB 4; Length 1327;
Best Local Similarity 30.4%; Pred. No. 1.5;
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

QY 6 VAAPFAAI-VYGSALAGVVPWGGGNGHNGGNSGDPSTLSIYQYGSANAALALQSDAR 64
DB 99 VAAAPVVPVAVSTSSAAGVAPNPAGSGNGNSPSSSSPTSS--SSSSPSSPGSSLAESPEAA 157
QY 65 KSETTIT---QSGYGNAGDVGGADNSTTIELTONG--FRNNATIDOWNAKNSDI 113
DB 158 GVSSTAPLPGCAAGPGTGVPVAGSALRELLAEACRGDVSVRKVLVDAAVNAKDM 212

RESULT 10
US-09-336-115C-6
; Sequence 6, Application US/09136115C
; Patent No. 6576244
; GENERAL INFORMATION:
; APPLICANT: Weltzin, Richard A.
; TITLE OF INVENTION: GUY, Bruno
; TITLE OF INVENTION: LT and CT in Parenteral Immunization
; FILE REFERENCE: 06132/055002
; CURRENT APPLICATION NUMBER: US/09/336.115C
; CURRENT FILING DATE: 1999-06-18
; PRIOR APPLICATION NUMBER: US 09/100,258
; PRIOR FILING DATE: 1998-06-19
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 745
; TYPE: PRT
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)...(20)
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 721
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-336-115C-6

Query Match      11.6%; Score 90.5; DB 4; Length 745;
Best Local Similarity 20.8%; Pred. No. 0.97;
Matches 40; Conservative 28; Mismatches 63; Indels 61; Gaps 8;

QY 8 APAAIYVSGSALAGVVPWQ-----GGGNGHNGGN----- 37
DB 102 AYCAVFLAINAAVGL---WNTIGVAVMGNGNGTSGPGSVIFNDPQGDSTQITCNRFE 158
QY 38 SSGPSTLSIYQYGSANAALALQSDARKSETTITQSGYVNG-----ADVQ 83
DB 159 STGPGKNSIDEFKKLNKAYQLIQALKNQSGFFELG--GNTKVSVVNYECROTADING 217
QY 84 G-----ADNSTIELTONGFRNNATIDOWNAKNSDITVGYGGNNAALVNYDQLVTRVWT 137
DB 218 GYQPCXKAKNGS--SSSSNGGSGSTQTATTATTDGVTITTYNNKATVKFD-----IT 270
QY 138 HEMAHANNATAN 149
DB 271 NNAEQLNQAN 282

RESULT 11
US-08-864-038A-3
; Sequence 3, Application US/08864038A
```

```
; Patent No. 6001592
; GENERAL INFORMATION:
; APPLICANT: Kunio NAKASHIMA et al.
; TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE CDNA, VECTOR
; TITLE OF INVENTION: CONTAINING SAID CDNA, HOST CELLS TRANSFORMED WITH SAID
; TITLE OF INVENTION: VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING
; TITLE OF INVENTION: SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBODY
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: 812-5 Hirano
; STREET: Isshinden
; CITY: Tsu-city
; STATE: Mie-prefecture
; COUNTRY: JAPAN
; ZIP: 514-01
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Microsoft Windows 95
; SOFTWARE: Word Perfect 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/864,038A
; FILING DATE: May 28, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-184459
; FILING DATE: 15-July-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: C. Bruce Hamburg
; REGISTRATION NUMBER: 22,389
; REFERENCE/DOCKET NUMBER: F-5610
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)986-2340
; TELEFAX: (212)953-7733
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 738
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORGANISM SOURCE:
; ORGANISM: Finctada fucata
; CELL TYPE: mantle epithelial cell
; FEATURE:
; NAME/KEY: peptide
; LOCATION: from 1 to 738
; IDENTIFICATION METHOD: E (by experiment)
US-08-864-038A-3

Query Match      11.5%; Score 89.5; DB 3; Length 738;
Best Local Similarity 35.4%; Pred. No. 1.2;
Matches 29; Conservative 4; Mismatches 26; Indels 23; Gaps 3;

QY 3 LKVAAPAAIYVSGSALAGVVPWGGGNGHNGGNSGDPSTLSIYQYGSANAALALQSD 62
DB 419 LKSSASASASASASAG-----GGGGGNGGGGGG-----GGGAGALA----- 460
QY 63 ARKSETTITQSGYGNAGDVGG 84
DB 461 -----AALAAAGAGGGLGGGG 477

RESULT 12
US-09-056-556-204
; Sequence 204, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
```

```

; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 204:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 943 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-09-056-556-204

Query Match 11.4%; Score 89; DB 4; Length 943;
Best Local Similarity 25.6%; Pred. No. 1.9;
Matches 33; Conservative 15; Mismatches 55; Indels 26; Gaps 5;

QY 12 IVVSGSALAGVVPQWGGGHNHGGNSGPDSTLSIYQVGSANAALALQSDAR---KSET 68
Db 159 IGLTSGLLGF-----GGLNSGTGN-----IGLFSGTGNGVIGNSGTGNWIGNSG 205

QY 69 TITQSGYNGADVQGGADNSTIELTQNGFRNNATIDQWNAKNSD---ITVGQY----- 118
Db 206 NSYNTGFNGSGDANTGFFNSGIANTGVGNAGNYTGSYNPGNSNTGGFNMGOYNTGYLNS 265

QY 119 GGNNAALVN 127
Db 266 GNYNTGLAN 274

RESULT 13
US-09-072-596-199
; Sequence 199, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dallon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Hedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; CORRESPONDENCE ADDRESS:
; NUMBER OF SEQUENCES: 350
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 199:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 943 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-09-072-596-199

Query Match 11.4%; Score 89; DB 4; Length 943;
Best Local Similarity 25.6%; Pred. No. 1.9;
Matches 33; Conservative 15; Mismatches 55; Indels 26; Gaps 5;

QY 12 IVVSGSALAGVVPQWGGGHNHGGNSGPDSTLSIYQVGSANAALALQSDAR---KSET 68
Db 159 IGLTSGLLGF-----GGLNSGTGN-----IGLFSGTGNGVIGNSGTGNWIGNSG 205

QY 69 TITQSGYNGADVQGGADNSTIELTQNGFRNNATIDQWNAKNSD---ITVGQY----- 118
Db 206 NSYNTGFNGSGDANTGFFNSGIANTGVGNAGNYTGSYNPGNSNTGGFNMGOYNTGYLNS 265

QY 119 GGNNAALVN 127
Db 266 GNYNTGLAN 274

RESULT 14
US-09-477-135A-131
; Sequence 131, Application US/09477135A
; Patent No. 6572865
; GENERAL INFORMATION:
; APPLICANT: Nano, Francis
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
; TITLE OF INVENTION: immunostimulatory Peptides
; FILE REFERENCE: 52888
; CURRENT APPLICATION NUMBER: US/09/477,135A
; CURRENT FILING DATE: 2000-01-03
; PRIOR APPLICATION NUMBER: 08990823
; PRIOR FILING DATE: 1997-12-15
; PRIOR APPLICATION NUMBER: US 96/10375
; PRIOR FILING DATE: 1996-06-14
; PRIOR APPLICATION NUMBER: 60/000,254
; PRIOR FILING DATE: 1995-06-15
; NUMBER OF SEQ ID NOS: 169
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 131
; LENGTH: 943
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; US-09-477-135A-131

Query Match 11.4%; Score 89; DB 4; Length 943;
Best Local Similarity 25.6%; Pred. No. 1.9;
Matches 33; Conservative 15; Mismatches 55; Indels 26; Gaps 5;

QY 12 IVVSGSALAGVVPQWGGGHNHGGNSGPDSTLSIYQVGSANAALALQSDAR---KSET 68
Db 266 IGLTSGLLGF-----GGLNSGTGN-----IGLFSGTGNGVIGNSGTGNWIGNSG 312

QY 69 TITQSGYNGADVQGGADNSTIELTQNGFRNNATIDQWNAKNSD---ITVGQY----- 118

```

Job time : 13 secs

Db 313 NSYNTGFGNSGDANTGFFNSGIANTGVGNAGNYNTGTSYNPNGNSNTGFGNMGQYNTGYLNS 372

QY 119 GGNNAALVN 127
Db 373 GNYNTGLAN 381

RESULT 15

US-09-072-967-204
; Sequence 204, Application US/09072967
; Patent No. 6592877
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skelky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 204:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 943 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-09-072-967-204

Query Match 11.4%; Score 89; DB 4; Length 943;
Best Local Similarity 25.6%; Pred. No. 1.9;
Matches 33; Conservative 15; Mismatches 55; Indels 26; Gaps 5;
QY 12 IVVSGSALAGVVPQWGGNHNHNGGNSGSPDSTLSIYQYGSANAALALQSDAR---KSET 68
Db 159 IGLTSGGLGF-----GGLNSGTGN-----IGLFNSGTGNVIGNSGTGNWIGNSG 205
QY 69 TTQSGYGNAGDVGGADNSTIELTQNGFRNNATIDOWNAKNSD---ITVQY----- 118
Db 206 NSYNTGFGNSGDANTGFFNSGIANTGVGNAGNYNTGTSYNPNGNSNTGFGNMGQYNTGYLNS 265
QY 119 GGNNAALVN 127
Db 266 GNYNTGLAN 274

Search completed: August 2, 2004, 14:58:31

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:54:48 ; Search time 36.8 Seconds

(without alignments)
1287.123 Million cell updates/sec

Title: US-09-543-407-12

Perfect score: 779
Sequence: 1 MKLKVAFAAIVVSGSALA.....VTRVTHMAHANNATANY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	517	66.4	151	12	US-09-741-873B-4
2	517	66.4	151	12	US-09-741-873B-4
3	439	56.4	131	12	US-09-741-873B-2
4	439	56.4	131	12	US-09-741-873B-2
5	100	12.8	445	15	US-10-369-493-20638
6	97.5	12.5	263	12	US-10-425-114-49960
7	94.5	12.1	678	12	US-10-282-122A-64573
8	94.5	12.1	688	14	US-10-032-585-7876
9	93	11.9	271	16	US-10-437-963-147343
10	92	11.8	573	9	US-09-841-835-8
11	92	11.8	949	9	US-09-841-835-10
12	92	11.8	1327	9	US-09-841-835-2
13	92	11.8	1327	10	US-09-972-115A-8
14	92	11.8	1327	14	US-10-199-937-4
15	91.5	11.7	705	16	US-10-437-963-203823

16	91	11.7	154	16	US-10-437-963-162284
17	90.5	11.6	745	8	US-08-834-666A-6
18	90.5	11.6	745	12	US-10-282-122A-58683
19	90	11.6	1778	14	US-10-238-075-749
20	89.5	11.5	270	16	US-10-437-963-122263
21	89.5	11.5	486	15	US-10-369-493-20619
22	89	11.4	943	9	US-09-996-634-131
23	89	11.4	943	10	US-09-997-182-131
24	89	11.4	943	10	US-09-997-181-131
25	89	11.4	943	14	US-10-193-002-199
26	89	11.4	943	14	US-10-084-843-204
27	89	11.4	3300	12	US-10-282-122A-64369
28	88.5	11.4	145	16	US-10-437-963-147748
29	88.5	11.4	191	16	US-10-437-963-105413
30	88	11.3	354	10	US-09-820-843A-21
31	88	11.3	616	16	US-10-437-963-193067
32	87.5	11.2	253	16	US-10-437-963-114193
33	87.5	11.2	628	12	US-10-282-122A-53269
34	87.5	11.2	892	10	US-09-952-287-5
35	87	11.2	482	14	US-10-156-761-8763
36	86	11.0	448	16	US-10-437-963-140685
37	86	11.0	2457	12	US-10-282-122A-49854
38	85.5	11.0	191	16	US-10-437-963-152921
39	85.5	11.0	942	12	US-10-282-122A-73345
40	85.5	11.0	1862	12	US-10-282-122A-49757
41	85	10.9	257	10	US-09-880-748-1494
42	85	10.9	257	12	US-10-293-418-1494
43	85	10.9	956	13	US-10-121-032-63
44	85	10.9	956	14	US-10-093-037-63
45	84.5	10.8	271	14	US-10-156-761-11721

ALIGNMENTS

RESULT 1

US-09-741-873B-4
; Sequence 4, Application US/09741873B
; Publication No. US20020081722A1
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; PRIOR FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patent version 3.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli

Query Match 66.4%; Score 517; DB 12; Length 151;
Best Local Similarity 66.4%; Pred. No. 1.7e+43;
Matches 100; Conservative 22; Mismatches 29; Indels 0; Gaps 0;
QY 1 MKLKVAFAAIVVSGSALAGVTFQGGGNGHNGGNSGPDSTLSIYQVGSANALALQ 60

Sequence 162284,
Sequence 6, Appli
Sequence 58683, A
Sequence 749, App
Sequence 122263, A
Sequence 20619, A
Sequence 131, App
Sequence 131, App
Sequence 131, App
Sequence 199, App
Sequence 204, App
Sequence 64369, A
Sequence 147748,
Sequence 105413,
Sequence 21, Appl
Sequence 193067,
Sequence 114193,
Sequence 53269, A
Sequence 5, Appli
Sequence 8763, Ap
Sequence 140685,
Sequence 49854, A
Sequence 152921,
Sequence 73345, A
Sequence 49757, A
Sequence 1494, Ap
Sequence 1494, Ap
Sequence 63, Appl
Sequence 63, Appl
Sequence 11721, A

```
Db 1 MKLLKVAALAAIVFSGSAGVVPQYGGGNGHGGGNSGPNSELNIYQYGGNSALALQ 60
QY 61 SDARKSETTITQSGYNGADVGCGADNSTIETQNGFRNATIDOWNAKNSDITVQGVGG 120
Db 61 TDARNSDLTITQGGGNGADVGCGSDSSIDLQTFQFGNSATLDQWNGKNSMTVKQFGG 120
QY 121 NNAALVNYDQVTRVVTHEMAHANNATANY 151
Db 121 GNGAAVDQTASNSVNVTVQVGFNNATAHQY 151
```

RESULT 2

```
US-09-741-873B-4
; Sequence 4, Application US/09741873B
; Publication No. US200400965A9
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; PRIOR FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
```

US-09-741-873B-4

```
Query Match 66.4%; Score 517; DB 12; Length 151;
Best Local Similarity 66.2%; Pred. No. 1.7e-43;
Matches 100; Conservative 22; Mismatches 29; Indels 0; Gaps 0;
APPLICANT: Normark, Staffan
TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
FILE REFERENCE: 012889-084
CURRENT APPLICATION NUMBER: US/09/741,873B
PRIOR FILING DATE: 2003-04-04
PRIOR APPLICATION NUMBER: SE 8801723-1
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: US 08/978,878
PRIOR FILING DATE: 1997-11-26
PRIOR APPLICATION NUMBER: US 07/347,189
PRIOR FILING DATE: 1989-05-04
PRIOR APPLICATION NUMBER: US 07/789,437
PRIOR FILING DATE: 1991-11-06
PRIOR APPLICATION NUMBER: US 07/970,846
PRIOR FILING DATE: 1992-11-03
PRIOR APPLICATION NUMBER: US 08/187,865
PRIOR FILING DATE: 1994-01-28
PRIOR APPLICATION NUMBER: US 08/318,519
PRIOR FILING DATE: 1994-10-05
NUMBER OF SEQ ID NOS: 10
SOFTWARE: PatentIn version 3.0
SEQ ID NO 4
LENGTH: 151
TYPE: PRT
ORGANISM: Escherichia coli
```

US-09-741-873B-4

```
QY 1 MKLLKVAALAAIVFSGSAGVVPQYGGGNGHGGGNSGPNSELNIYQYGSANAALALQ 60
Db 1 MKLLKVAALAAIVFSGSAGVVPQYGGGNGHGGGNSGPNSELNIYQYGGNSALALQ 60
QY 61 SDARKSETTITQSGYNGADVGCGADNSTIETQNGFRNATIDOWNAKNSDITVQGVGG 120
Db 61 TDARNSDLTITQGGGNGADVGCGSDSSIDLQTFQFGNSATLDQWNGKNSMTVKQFGG 120
QY 121 NNAALVNYDQVTRVVTHEMAHANNATANY 151
Db 121 GNGAAVDQTASNSVNVTVQVGFNNATAHQY 151
```

RESULT 3

```
US-09-741-873B-2
; Sequence 2, Application US/09741873B
; Publication No. US20020081722A1
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; PRIOR FILING DATE: 2003-04-04
```

```
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-2
```

```
Query Match 56.4%; Score 439; DB 12; Length 131;
Best Local Similarity 62.6%; Pred. No. 8e-36;
Matches 82; Conservative 21; Mismatches 28; Indels 0; Gaps 0;
QY 21 GVTPQWGGGNGHGGGNSGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGAD 80
Db 1 GVTPQWGGGNGHGGGNSGPNSELNTYQYGGNSALALQTDARNSDLTITQGGGNGAD 60
QY 81 VGGQADNSTIETQNGFRNATIDOWNAKNSDITVQYGGNAALVNYDQVTRVVTHEM 140
Db 61 VGGQSDSSIDLQTFQFGNSATLDQWNGKNSMTVKQFGGNGAAVDQTASNSVNVTVQ 120
QY 141 AHANNATANY 151
Db 121 GFGNNATAHQY 131
```

RESULT 4

```
US-09-741-873B-2
; Sequence 2, Application US/09741873B
; Publication No. US200400965A9
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; CURRENT FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-2
```

Query Match 56.4%; Score 439; DB 12; Length 131;


```

; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 64573
; LENGTH: 678
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-10-282-122A-64573

Query Match      12.1%; Score 94.5; DB 12; Length 678;
Best Local Similarity 25.7%; Pred. No. 1.1;
Matches 39; Conservative 24; Mismatches 48; Indels 41; Gaps 8;

QY 15 SCSALAGVVPWGGGNNHNGGSGDPDSTLSIYQYGSANAALALQ--SDARKSETTITQ 72
Db 384 SGGNGLG----FGSGNGNIGFFNSG--NNNIGMNGSGNGVGALSVEFGSSAERS----- 432
QY 73 SGYNGADVQGGADNS-----TIELTQNGFRNNATIDQ--WNAKNSDITVGVGGNN 122
Db 433 SGFGNSGELSTGLGNSGQLSTGWFNSATTTSTGWFNSGTTNTGWFNSGTTNTGIGNSGGN- 491
QY 123 AALVNYDQLVTRVVTHEM-----AHANNATAN 149
Db 492 -----LVTGSMGLFNSGHTNTGSPN 511

RESULT 8
US-10-032-585-7876
; Sequence 7876, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
; APPLICANT: Terry, Roemer D.
; APPLICANT: Bo, Jiang
; APPLICANT: Charles, Boone
; APPLICANT: Howard, Bussey
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7876
; LENGTH: 688
; TYPE: PRT
; ORGANISM: Candida albicans
US-10-032-585-7876

Query Match      12.1%; Score 94.5; DB 14; Length 688;
Best Local Similarity 25.8%; Pred. No. 1.1;
Matches 31; Conservative 24; Mismatches 42; Indels 23; Gaps 4;

QY 30 GNHNGGSGSGDPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGADVQGGADNST 89
Db 488 GNNNGSGNSGTTNNNNYNN-----KSISKNE-----IDGDDLNTSIINN 531
QY 90 IELTQNGFRNNATIDQWNAKNSDITVGVGGNNAALVNYDQLVTRVVTHEMAHANNATAN 149
Db 532 TGLTNN---NNSKSPAKSKKSNFD-----NNSNSALNLDKSKLKINTNEITN-SETTSN 584

RESULT 9
US-10-437-963-147343
; Sequence 147343, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

```

```

; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 147343
; LENGTH: 271
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_47881C.1.pap
US-10-437-963-147343

Query Match      11.9%; Score 93; DB 16; Length 271;
Best Local Similarity 22.2%; Pred. No. 0.49;
Matches 34; Conservative 18; Mismatches 59; Indels 42; Gaps 5;

QY 14 VSGSALAGVVPWGGG--GNHNGGSGSGDPDSTLSIYQYGSANAALALQSDARKSETTITQ 72
Db 7 VVGRAAAAAYPEVSGSGDGGSGGGGEGGE-----GGDGSVAAVNPEAGSGGDRSS 58
QY 73 SGYNGADVQGGADNSTIELTQNGFRNNATIDQNA-----KNSDITVGV 118
Db 59 GEGGSGSGGLGRQR-----YWNRLSTERQRLVDHVFKNSDVVCDFV 103

QY 119 GGNAALVNYDQLVTRVVTHEMAHANNATANQY 151
Db 104 SGVGPIS----AARKVYVYANDLNPITAVEY 132

RESULT 10
US-09-841-835-8
; Sequence 8, Application US/09841835
; Patent No. US20020076795A1
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; TITLE OF INVENTION: OF USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/841,835
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/196,387
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 673 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-841-835-8

```

```

Query Match      11.8%; Score 92; DB 9; Length 673;
Best Local Similarity 30.4%; Pred. No. 1.9;
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

QY      6 VAAFAAI-VVSGALACVPPQWGGGHHNGGSSGPDSTLSIYQYGSANAALQSDAR 64
      ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db      99 VAAAPVPAVSTSAAGVARNPAGSGNNPSSSSSTSS--SSSPSPSGSSLAESPAA 157

QY      65 KSETTIT----QSGYNGNADVGGQADNSTIETLQNG--FRNNATIDOWNAKNSDI 113
      : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db      158 GVSTAPLPGCAAGPGTGVPAVSGALRELLFACRNGDVSRVKKRILVDAAVNAKDM 212

```

RESULT 11
US-09-841-835-10
; Sequence 10, Application US/09841835
; Patent No. US20020076795A1
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; TITLE OF INVENTION: OF USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/841/835

```

Query Match      11.8%; Score 92; DB 9; Length 949;
Best Local Similarity 30.4%; Pred. No. 2.9;
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

Qy 6 VAAPAAI-VVGSALACVFPQWGGNHGGNGSSGPDSTLSIYQYGSAAALALQSDAR 64
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 99 VAAPVVPVYSTSSAAGVAPNPAGSGNGNSPSSSSPTSS-SSSSPSPSSSLAESFEAA 157
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :

Qy 65 KSETTIT---QSGYNGNGADVGGADNSTIETLQTNG--FRNNATIDOWNAKNSDI 113
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 158 GVSSTAPLGFGAAGPGTGVPAVSGALRELLAEACFNGDVS RVKRLVDAAVNAKDM 212
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :

RESULT 12

```

[illegible]

```
; PRIOR FILING DATE: 2000-04-10
; PRIOR APPLICATION NUMBER: US 60/129,123
; PRIOR FILING DATE: 1999-04-13
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 1327
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-972-115A-8

Query Match
  11.8%; Score 92; DB 10; Length 1327;
Best Local Similarity 30.4%; Pred. No. 4.4;
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

QY 6 VAAFAAI-VVSGSALAGVVPQWGGGNNHNGGNSGDPSTLSIYQYGSANAALALQSDAR 64
Db 99 VAAAPVVPVAVSTSSAAGVAPNAPAGSGNNPSSSSPTSS-SSSSPSPGSSLAESPEAA 157
QY 65 KSETTIT---QSGYNGADVCGGADNSTIETLQNG--FRNNATIDOWNAKNSDI 113
Db 158 GVSSTAPLPGGAAGPGTGVPVAVSGALRELLEACRNGDVSRVKRLVDAANVNAKDM 212

RESULT 14
US-10-199-937-4
; Sequence 4, Application US/10199937
; Publication No. US20030190739A1
; GENERAL INFORMATION:
; APPLICANT: Christenson, Erik
; APPLICANT: Demaggio, Anthony J.
; APPLICANT: Goldman, Phyllis S.
; APPLICANT: McElligott, David L.
; TITLE OF INVENTION: TANKYRASE2 MATERIALS AND METHODS
; FILE REFERENCE: 27866/36559
; CURRENT APPLICATION NUMBER: US/10/199,937
; CURRENT FILING DATE: 2002-07-22
; PRIOR APPLICATION NUMBER: US/09/606,035
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 60/141,582
; PRIOR FILING DATE: 1999-06-29
; NUMBER OF SEQ ID NOS: 178
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1327
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-199-937-4

Query Match
  11.8%; Score 92; DB 14; Length 1327;
Best Local Similarity 30.4%; Pred. No. 4.4;
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

QY 6 VAAFAAI-VVSGSALAGVVPQWGGGNNHNGGNSGDPSTLSIYQYGSANAALALQSDAR 64
Db 99 VAAAPVVPVAVSTSSAAGVAPNAPAGSGNNPSSSSPTSS-SSSSPSPGSSLAESPEAA 157
QY 65 KSETTIT---QSGYNGADVCGGADNSTIETLQNG--FRNNATIDOWNAKNSDI 113
Db 158 GVSSTAPLPGGAAGPGTGVPVAVSGALRELLEACRNGDVSRVKRLVDAANVNAKDM 212

RESULT 15
US-10-437-963-203823
; Sequence 203823, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
```

```
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 203823
; LENGTH: 705
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_98970C.1.pap
US-10-437-963-203823

Query Match
  11.7%; Score 91.5; DB 16; Length 705;
Best Local Similarity 24.7%; Pred. No. 2.3;
Matches 42; Conservative 19; Mismatches 56; Indels 53; Gaps 7;

QY 12 IVVSGSALAGVVPQWGGGNNHNGGNSGDPSTLSIYQYGSANAALALQSDARKSET 68
Db 112 IFSGGSSSAGS---GGGGGGGGSTASSPLTNALPTGNICPSARVASAAPPRRARP 167
QY 69 TITQSG---YNGG-----ADVGGADNSTIETLQNGER 98
Db 168 DVLGSGTGHYGHGSIIRGGGGMTTPRSSIDASPHYHSYRSRPAQSGSGGLQEVTRAG-- 225
QY 99 NNATIDOWNAKNSDIIVGOYGGNNAALVNYDOLVTRVVTHEMAHANNATA 148
Db 226 -----NEWYKK-----GHYG---EALRYDQVALCPDPSAACSNNRAAA 261

Search completed: August 2, 2004, 15:36:09
Job time : 37.8 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:48:33 ; Search time 167.9 Seconds
(without alignments)
877.809 Million cell updates/sec

Title: US-09-543-407-12
Perfect score: 779
Sequence: 1 MKLLKVAFAAIVVGSALA.....VTRVTHMAHANNATANYQ 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 6019581 seqs, 976053577 residues

Total number of hits satisfying chosen parameters: 6019581

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_AA_Main.*

1: /cgn2_6/ptodata/2/paa/PCTUS_COMB.pcp.*
2: /cgn2_6/ptodata/2/paa/US06_COMB.pcp.*
3: /cgn2_6/ptodata/2/paa/US07_COMB.pcp.*
4: /cgn2_6/ptodata/2/paa/US08_COMB.pcp.*
5: /cgn2_6/ptodata/2/paa/US081_COMB.pcp.*
6: /cgn2_6/ptodata/2/paa/US082_COMB.pcp.*
7: /cgn2_6/ptodata/2/paa/US083_COMB.pcp.*
8: /cgn2_6/ptodata/2/paa/US084_COMB.pcp.*
9: /cgn2_6/ptodata/2/paa/US085_COMB.pcp.*
10: /cgn2_6/ptodata/2/paa/US086_COMB.pcp.*
11: /cgn2_6/ptodata/2/paa/US087_COMB.pcp.*
12: /cgn2_6/ptodata/2/paa/US088_COMB.pcp.*
13: /cgn2_6/ptodata/2/paa/US089_COMB.pcp.*
14: /cgn2_6/ptodata/2/paa/US090_COMB.pcp.*
15: /cgn2_6/ptodata/2/paa/US091_COMB.pcp.*
16: /cgn2_6/ptodata/2/paa/US092_COMB.pcp.*
17: /cgn2_6/ptodata/2/paa/US093_COMB.pcp.*
18: /cgn2_6/ptodata/2/paa/US094_COMB.pcp.*
19: /cgn2_6/ptodata/2/paa/US095_COMB.pcp.*
20: /cgn2_6/ptodata/2/paa/US096_COMB.pcp.*
21: /cgn2_6/ptodata/2/paa/US097A_COMB.pcp.*
22: /cgn2_6/ptodata/2/paa/US097B_COMB.pcp.*
23: /cgn2_6/ptodata/2/paa/US098_COMB.pcp.*
24: /cgn2_6/ptodata/2/paa/US099A_COMB.pcp.*
25: /cgn2_6/ptodata/2/paa/US099B_COMB.pcp.*
26: /cgn2_6/ptodata/2/paa/US100_COMB.pcp.*
27: /cgn2_6/ptodata/2/paa/US101_COMB.pcp.*
28: /cgn2_6/ptodata/2/paa/US102_COMB.pcp.*
29: /cgn2_6/ptodata/2/paa/US103_COMB.pcp.*
30: /cgn2_6/ptodata/2/paa/US104_COMB.pcp.*
31: /cgn2_6/ptodata/2/paa/US106_COMB.pcp.*
32: /cgn2_6/ptodata/2/paa/US107_COMB.pcp.*
33: /cgn2_6/ptodata/2/paa/US60_COMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result Query
No. Score Match Length DB ID Description

1	779	100.0	151	19	US-09-543-407-12	Sequence 12, Appl
2	696	89.3	151	19	US-09-543-407-14	Sequence 14, Appl
3	690	89.2	151	19	US-09-543-407-5	Sequence 5, Appl
4	690	88.6	151	6	US-08-233-642A-57	Sequence 57, Appl
5	655	84.1	151	19	US-09-543-407-24	Sequence 24, Appl
6	614	78.8	151	19	US-09-543-407-26	Sequence 26, Appl
7	613	78.7	151	19	US-09-543-407-18	Sequence 18, Appl
8	611	78.4	151	19	US-09-543-407-20	Sequence 20, Appl
9	608	78.0	131	19	US-09-543-407-31	Sequence 31, Appl
10	605	77.7	151	19	US-09-543-407-28	Sequence 28, Appl
11	604	77.5	151	19	US-09-543-407-22	Sequence 22, Appl
12	603	77.4	151	19	US-09-543-407-30	Sequence 30, Appl
13	580	74.5	151	19	US-09-543-407-16	Sequence 16, Appl
14	560	71.9	120	6	US-08-233-642A-55	Sequence 55, Appl
15	520	66.8	151	19	US-09-543-407-7	Sequence 7, Appl
16	517	66.4	151	13	US-08-978-878-4	Sequence 4, Appl
17	517	66.4	151	21	US-09-741-873B-4	Sequence 4, Appl
18	515	66.1	151	33	US-60-352-946-2	Sequence 2, Appl
19	515	66.1	151	33	US-60-444-371-2	Sequence 2, Appl
20	481	61.7	158	16	US-09-252-691-5834	Sequence 5834, Ap
21	481	61.7	158	16	US-09-252-691C-5834	Sequence 5834, Ap
22	481	61.7	158	30	US-10-417-886-5834	Sequence 5834, Ap
23	476	61.1	109	19	US-09-543-407-34	Sequence 34, Appl
24	439	56.4	131	13	US-08-978-878-2	Sequence 2, Appl
25	439	56.4	131	21	US-09-741-873B-2	Sequence 2, Appl
26	347	44.5	68	19	US-09-543-407-37	Sequence 37, Appl
27	335	43.0	109	19	US-09-543-407-35	Sequence 35, Appl
28	247.5	31.8	70	19	US-09-543-407-32	Sequence 32, Appl
29	237	30.4	48	19	US-09-543-407-39	Sequence 39, Appl
30	100.5	12.9	151	19	US-09-543-407-6	Sequence 6, Appl
31	100.5	12.9	186	16	US-09-252-691C-5833	Sequence 5833, Ap
32	100.5	12.9	186	16	US-09-252-691C-5833	Sequence 5833, Ap
33	100.5	12.9	186	30	US-10-417-886-5833	Sequence 5833, Ap
34	100	12.8	445	29	US-10-369-493-20638	Sequence 20638, A
35	100	12.8	445	33	US-60-360-039-20638	Sequence 20638, A
36	100	12.8	1249	30	US-10-455-719-358	Sequence 358, App
37	100	12.8	1249	33	US-60-446-775-358	Sequence 358, App
38	100	12.8	1249	33	US-60-446-775-358	Sequence 358, App
39	100	12.8	1266	27	US-10-179-131-9317	Sequence 9317, Ap
40	98.5	12.6	850	20	US-09-614-150-24084	Sequence 24084, A
41	98.5	12.6	850	20	US-09-614-150A-24084	Sequence 24084, A
42	98.5	12.6	850	33	US-60-151-637-24179	Sequence 24179, A
43	98.5	12.6	850	33	US-60-191-681-19019	Sequence 19019, A
44	98.5	12.6	1028	20	US-09-614-150-14916	Sequence 14916, A
45	98.5	12.6	1028	20	US-09-614-150A-14916	Sequence 14916, A

ALIGNMENTS

RESULT 1
US-09-543-407-12
; Sequence 12, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043 406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.

```

US-09-543-407-12
Query Match      100.0%; Score 779; DB 19; Length 151;
Best Local Similarity 100.0%; Pred. No. 2.9e-74;
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGGNGSGPDSLTLSIYQGSANAALALQ 60
DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGGNGSGPDSLTLSIYQGSANAALALQ 60

QY 61 SPARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120
DB 61 SPARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120

QY 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151
DB 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151

RESULT 2
US-09-543-407-14
; Sequence 14, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-14

Query Match      89.3%; Score 696; DB 19; Length 151;
Best Local Similarity 87.6%; Pred. No. 2e-65;
Matches 141; Conservative 0; Mismatches 0; Indels 20; Gaps 2;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGGNGSGPDSLTLSIYQGSANAALALQ 60
DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGGNGSGPDSLTLSIYQGSANAALALQ 60

QY 61 SPARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120
DB 61 SPARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 117

QY 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151
DB 118 -----YDQLVTRVVTHEMAHASVMVQVGFNGNATANQY 151

RESULT 3
US-09-543-407-5
; Sequence 5, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Salmonella enteritidis
US-09-543-407-5

Query Match      89.2%; Score 695; DB 19; Length 151;
Best Local Similarity 90.1%; Pred. No. 2.5e-65;
Matches 136; Conservative 4; Mismatches 11; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGGNGSGPDSLTLSIYQGSANAALALQ 60
DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGGNGSGPDSLTLSIYQGSANAALALQ 60

QY 61 SPARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120
DB 61 SPARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120

QY 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151
DB 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151

RESULT 4
US-08-233-642A-57
; Sequence 57, Application US/08233642A
; GENERAL INFORMATION:
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Clouthier, Sharon C.
; APPLICANT: Doran, James L.
; APPLICANT: Doran, James L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
; TITLE OF INVENTION: BASED VACCINES
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESS: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEDANBERRY
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 151 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-233-642A-57

Query Match      88.6%; Score 690; DB 6; Length 151;
Best Local Similarity 89.4%; Pred. No. 8.6e-65;
Matches 135; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

```



```
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGDPSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGDPSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGCGADNSTIETQNGFRNNATIDQWNAKNSDITVGOYGG 120
DB 61 SDARKSETTITQSGYNGADVGCGADNSTIETQNGFRNNATIDQWNAKNSDITVGOYGG 120
QY 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151
DB 121 NNPALVNTASDSSVMVRQVGFGNNTANQY 151

RESULT 5
US-09-543-407-24
; Sequence 24, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-24
Query Match 84.1%; Score 655; DB 19; Length 151;
Best Local Similarity 79.8%; Pred. No. 4.6e-61;
Matches 134; Conservative 0; Mismatches 0; Indels 34; Gaps 2;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGDPSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGDPSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGCGADNSTIETQNGFRNNATIDQWNAKNSDITVGOYGG 120
DB 61 SDARKSETTITQSGYNGADVGCGADNSTIETQNGFRNNATIDQWNAKNSDITVGOYGG 109
QY 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151
DB 110 -----NYDQLVTRVVTHEMAHANNATANQY 151

RESULT 6
US-09-543-407-26
; Sequence 26, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 151
```

```
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-26
Query Match 78.8%; Score 614; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 1.1e-56;
Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGDPSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGDPSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGCGADNSTIETQNGFRNNATIDQWNAKNSDITVGOYGG 120
DB 61 LVTRVVTHEMAHAGYNGADVGCGADNSTIETQNGFRNNATIDQWNAKNSDITVGOYGG 120
QY 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151
DB 121 NNAALVNTASDSSVMVRQVGFGNNTANQY 151

RESULT 7
US-09-543-407-18
; Sequence 18, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-18
Query Match 78.7%; Score 613; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 1.4e-56;
Matches 122; Conservative 7; Mismatches 22; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGDPSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGDPSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGCGADNSTIETQNGFRNNATIDQWNAKNSDITVGOYGG 120
DB 61 SDARKSETTITQSGYNGADVGCGADNSTIETQNGFRNNATIDQWNAKNSDITVGOYGG 120
QY 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151
DB 121 NNAALVNTASDSSVMVRQVGFGNNTANQY 151

RESULT 8
US-09-543-407-20
; Sequence 20, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
```

```
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-20

Query Match          78.4%; Score 611; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 2.2e-56;
Matches 122; Conservative 7; Mismatches 22; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSSGGPSTLSIYQYGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSSGGPSTLSIYQYGSANAALALQ 60

QY 61 SPARKSETTITQSGYNGADVCGADNSTIELTQNGFRNNATIDOWNAKNSDITVQYGG 120
Db 61 SPARKYDQLVTRVVTHEMAHAGQAGADNSTIELTQNGFRNNATIDOWNAKNSDITVQYGG 120

QY 121 NNAALVNVQLVTRVVTHEMAHANNATANOY 151
Db 121 NNAALVNVQTASDSSVMVRQVGFNNATANOY 151

RESULT 9
US-09-543-407-31
; Sequence 31, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Salmonella enteritidis
US-09-543-407-31

Query Match          78.0%; Score 608; DB 19; Length 131;
Best Local Similarity 88.5%; Pred. No. 3.9e-56;
Matches 116; Conservative 4; Mismatches 11; Indels 0; Gaps 0;

QY 21 GVVPQWGGGNGGNGSSGGPSTLSIYQYGSANAALALQSDARKSETTITQSGYNGAD 80
Db 1 GVVPQWGGGNGGNGSSGGPSTLSIYQYGSANAALALQSDARKSETTITQSGYNGAD 60

QY 81 VQGGADNSTIELTQNGFRNNATIDOWNAKNSDITVQYGGNNAALVNYDQLVTRVVTHEM 140
Db 61 VQGGADNSTIELTQNGFRNNATIDOWNAKNSDITVQYGGNNAALVNTASDSSVMVRQV 120

QY 141 AHANNATANOY 151
Db 121 GFGNNATANOY 131
```

```
RESULT 10
US-09-543-407-28
; Sequence 28, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-28

Query Match          77.7%; Score 605; DB 19; Length 151;
Best Local Similarity 80.1%; Pred. No. 9.8e-56;
Matches 121; Conservative 5; Mismatches 25; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSSGGPSTLSIYQYGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSSGGPSTLSIYQYGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGADVCGADNSTIELTQNGFRNNATIDOWNAKNSDITVQYGG 120
Db 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFNNATIDOWNAKNSDITVQYGG 120

QY 121 NNAALVNVQLVTRVVTHEMAHANNATANOY 151
Db 121 NNAALVNVQTASDSSVMVRQVGFNNATANOY 151

RESULT 11
US-09-543-407-22
; Sequence 22, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 22
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-22

Query Match          77.5%; Score 604; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 1.2e-55;
Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSSGGPSTLSIYQYGSANAALALQ 60
```

Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSSGPDSTLSIYQYGSANAALQ 60
Qy 61 SDARKSETTTTQSGYNGADVGQADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120
Db 61 SDARKSETTTTQSGYNGADVGQADNYDLVTRVVTHEMAHADQWNAKNSDITVGYGG 120
Qy 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151
Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 12
US-09-543-407-30
; Sequence 30, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543.407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-30

Query Match 77.4%; Score 603; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 1.6e-55;
Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;
Qy 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSSGPDSTLSIYQYGSANAALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSSGPDSTLSIYQYGSANAALQ 60
Qy 61 SDARKSETTTTQSGYNGADVGQADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120
Db 61 SDARKSETTTTQSGYNGADVGQADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120
Qy 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151
Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 13
US-09-543-407-16
; Sequence 16, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543.407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence

; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-16
Query Match 74.5%; Score 580; DB 19; Length 151;
Best Local Similarity 80.1%; Pred. No. 4.5e-53;
Matches 121; Conservative 7; Mismatches 23; Indels 0; Gaps 0;
Qy 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSSGPDSTLSIYQYGSANAALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVYDQVTRVVTHEMAHASGPDSTLSIYQYGSANAALQ 60
Qy 61 SDARKSETTTTQSGYNGADVGQADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120
Db 61 SDARKSETTTTQSGYNGADVGQADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120
Qy 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151
Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 14
US-08-233-642A-55
; Sequence 55, Application US/08233642A
; GENERAL INFORMATION:
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Clouthier, Sharon C.
; APPLICANT: Doran, James L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
; TITLE OF INVENTION: BASED VACCINES
; NUMBER OF SEQUENCES: 58
; NUMBER OF SEQUENCES: -
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,642A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 55:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 120 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-233-642A-55

Query Match 71.9%; Score 560; DB 6; Length 120;
Best Local Similarity 100.0%; Pred. No. 4.5e-51;
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 22 VVPQWGGGNGHNGSSGPDSTLSIYQYGSANAALQSDARKSETTTTQSGYNGADV 81

Db 1 VVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGADV 60

QY 82 GQADNSTIELTQNGFRNNATIDOWNAKNSDITVGYGNNAAALVN 127

Db 61 GQADNSTIELTQNGFRNNATIDOWNAKNSDITVGYGNNAAALVN 106

RESULT 15

US-09-543-407-7
; Sequence 7, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-543-407-7

Query Match 66.8%; Score 520; DB 19; Length 151;
Best Local Similarity 66.9%; Pred. No. 1,1e-46;
Matches 101; Conservative 21; Mismatches 29; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60

Db 1 MKLLKVAATAAIVFSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGADVGGADNSTIELTQNGFRNNATIDOWNAKNSDITVGYGG 120

Db 61 TDARNSDLTITQGGGNGADVQGGSDSDSIDLTQGFNGSATLDOWNGKNSMTVKQFGG 120

QY 121 NNAALVNDQVTRVVTHEMAHANNATANQY 151

Db 121 GNGAAVDQTAASNSVNVTVQVFGNNATAHQY 151

Search completed: August 2, 2004, 15:26:42
Job time : 168.9 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:49:38 ; Search time 17.8 Seconds
(without alignments)
888.146 Million cell updates/sec

Title: US-09-543-407-12

Perfect score: 779
Sequence: 1 MKLLKVAAPAAIVSGSALA.....VTRVVTHEMAHANNATANQY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 601315 seqs, 104695340 residues

Total number of hits satisfying chosen parameters: 601315

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents_AA_New:*
1: /cgn2_6/prodata/2/paa/PCT_NEW_COMB.pcp:*
2: /cgn2_6/prodata/2/paa/US05_NEW_COMB.pcp:*
3: /cgn2_6/prodata/2/paa/US07_NEW_COMB.pcp:*
4: /cgn2_6/prodata/2/paa/US08_NEW_COMB.pcp:*
5: /cgn2_6/prodata/2/paa/US09_NEW_COMB.pcp:*
6: /cgn2_6/prodata/2/paa/US10_NEW_COMB.pcp:*
7: /cgn2_6/prodata/2/paa/US60_NEW_COMB.pcp:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	517	66.4	151	5	US-09-741-873C-4
2	439	56.4	131	5	US-09-741-873C-2
3	97.5	12.5	258	6	US-10-425-115-300390
4	97.5	12.5	295	6	US-10-425-115-312488
5	94.5	12.1	719	5	US-09-248-796A-17559
6	92	11.8	299	6	US-10-170-205E-35751
7	92	11.8	1203	6	US-10-170-205E-741
8	92	11.8	1327	1	PCT-US04-02338-49
9	91.5	11.7	256	6	US-10-425-115-301334
10	91.5	11.7	443	6	US-10-100-683-7608
11	91.5	11.7	841	7	US-60-565-632-7906
12	91.5	11.7	841	7	US-60-579-062-7906
13	90.5	11.6	234	6	US-10-767-701-45603
14	90	11.6	586	1	PCT-US03-24982A-317
15	89	11.3	573	7	US-60-565-632-7907
16	88	11.3	573	7	US-60-579-062-7907
17	87.5	11.2	179	6	US-10-425-115-346132
18	87.5	11.2	892	5	US-09-952-267B-5
19	87.5	11.2	892	6	US-10-872-768-5
20	87.5	11.2	892	6	US-10-872-768-5
21	86	11.0	412	7	US-60-579-062-7905
22	86	11.0	412	7	US-60-579-062-7905
23	85	11.0	520	6	US-10-479-638-21
24	85	10.9	956	6	US-10-093-037A-63
25	84.5	10.8	251	6	US-10-854-439-256
26	83.5	10.7	251	6	US-10-854-439-249

27 82.5 10.6 873 5 US-09-952-267B-13 Sequence 13, Appl
28 82.5 10.6 873 6 US-10-872-768-13 Sequence 13, Appl
29 82.5 10.6 873 6 US-10-872-768-13 Sequence 13, Appl
30 82 10.5 193 6 US-10-425-115-254240 Sequence 23240, A
31 82 10.5 386 6 US-10-100-683-10326 Sequence 10326, A
32 82 10.5 386 6 US-10-100-683-10327 Sequence 10327, A
33 82 10.5 386 6 US-10-798-512-100 Sequence 100, App
34 82 10.5 555 1 PCT-US04-13211-3 Sequence 3, Appl
35 82 10.5 555 6 US-10-835-904-3 Sequence 3, Appl
36 82 10.5 1627 6 US-10-170-205E-16659 Sequence 16659, A
37 82 10.5 1905 1 PCT-US04-09388-9 Sequence 9, Appl
38 81.5 10.5 376 6 US-10-491-733-2 Sequence 2, Appl
39 81.5 10.5 508 6 US-10-425-115-285216 Sequence 285216, A
40 81.5 10.5 532 6 US-10-425-115-285214 Sequence 285214, A
41 81.5 10.5 1871 1 PCT-US03-02038-26 Sequence 26, Appl
42 81.5 10.5 4249 1 PCT-US04-04300-4 Sequence 4, Appl
43 81 10.4 131 7 US-60-565-632-11109 Sequence 11109, A
44 81 10.4 131 7 US-60-579-062-11109 Sequence 11109, A
45 81 10.4 201 6 US-10-425-115-309662 Sequence 309662, A

ALIGNMENTS

RESULT 1

US-09-741-873C-4
; Sequence 4 Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873C-4

Query Match 66.4%; Score 517; DB 5; Length 151;
Best Local Similarity 66.2%; Pred. No. 1.1e-36;
Matches 100; Conservative 22; Mismatches 29; Indels 0; Gaps 0;
QY 1 MKLLKVAAPAAIVSGSALAGVVPQWGGGNGGNGSGPDLSTLYQVGSANAALQ 60
Db 1 MKLLKVAAPAAIVFGSSAVAGVVPQYGGGNGGNGSGNSLNLYQYGGNSALALQ 60
QY 61 SPARSETTITOSGGNGADYGGQADNSTIELTQNGFNATIDOWNAKSDITVGOYGG 120
Db 61 TPARNSDITITQGGNGADYGGQSDSSIDLTRGFGNSALDQWNGKSENIVKQFGG 120
QY 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151
Db 121 GNGAAVDQTASNSVNVTVQVGFNNTAQAQY 151

```
RESULT 2
US-09-741-873C-2
; Sequence 2, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012899-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8501723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 2
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873C-2

Query Match          56.4%; Score 439; DB 5; Length 131;
Best Local Similarity 62.6%; Pred. No. 3.9e-30;
Matches 82; Conservative 21; Mismatches 28; Indels 0; Gaps 0;

QY 21 GVVPQGGGNNHGGNSGPDSTLSIYQGSANAALQSDARKSETTITQSGYNGAD 80
Db 1 GVVPQGGGNNHGGNSGPNSELNTYQGGNSALALQTDARNSLTITQHGCGNGAD 60

QY 81 VGGGADNSTIELTQGFNRNATIDQNNAKNSDITVQYGGNNAALVNDQLVTRVTHEM 140
Db 61 VGGSDSSIDITQRFNGSNALTDQNGKNSMTVKQFGGNGRAVDQTASNSVNTQV 120

QY 141 AHANNATANQY 151
Db 121 GFGNNATAHQY 131

RESULT 3
US-10-425-115-300390
; Sequence 300390, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 300390
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(258)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_37025C.1.pap
```

US-10-425-115-300390

```
Query Match          12.5%; Score 97.5; DB 6; Length 258;
Best Local Similarity 26.6%; Pred. No. 0.86;
Matches 34; Conservative 18; Mismatches 53; Indels 23; Gaps 5;

QY 17 SALAGVVPQWCGGNNHGGNSGPDSTLSIYQGSANAALQSDARKSETTITQSG-- 74
Db 78 SSIAG-----GGGGQGGGGTNGSGSGGGSGYGSSTAAASGPGSSGNVADAEGKGAG 132

QY 75 -----YNGA--DVGGADNSTIEL--TQGFNRNATIDQNNAKNSDITVQYGG 120
Db 133 GCMGGGANGAYGSGAGGKGVKGEVSGVALAPNSGYNGAADATGGGSG--AGGHHG 190

QY 121 NNAALVNY 128
Db 191 GAAGAPSY 198
```

RESULT 4

```
US-10-425-115-312468
; Sequence 312468, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 312468
; LENGTH: 295
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_48027C.1.pap
US-10-425-115-312468
```

```
Query Match          12.5%; Score 97.5; DB 6; Length 295;
Best Local Similarity 29.7%; Pred. No. 1;
Matches 38; Conservative 15; Mismatches 50; Indels 25; Gaps 4;

QY 26 WGGGNNHGGNSGPDSTLSIYQGSANAALQSDARKSETTITQSGYNGADVGGA 85
Db 138 YGGGYSSGGYSSG-GYAAAGYGVSGSGNYSNAGGGYSGS-----DYGNGAASGVA 192

QY 86 DNSTIELTQGFNRNATIDQNNAKNSDITVQYGGNNAAL 125
Db 193 NNLSSGYSNGRYNTIGSSDNGTGGYNSYPNPGAGNTYNTGGSSSGGTGTFGFGGGGFG 252

QY 126 VNYDQLVT 133
Db 253 NNFAGNVT 260
```

RESULT 5

```
US-09-248-796A-17559
; Sequence 17559, Application US/09248796A
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
```

; SEQ ID NO 17559
; LENGTH: 719
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-17559

Query Match 12.1%; Score 94.5; DB 5; Length 719;
Best Local Similarity 25.8%; Pred. No. 4.9;
Matches 31; Conservative 24; Mismatches 42; Indels 23; Gaps 4;

QY 30 GNHNGGNSGPDSTLSIYQGSANALALOSDARKSETTITQSGYNGADVGQGDNST 89
DB 519 GNNNGSGNSTTNNNNYNN-----KSISKQNE-----IDGDDLPNTSTINN 562
QY 90 IELTQNGFRNATIDQNAKSDITVGOYGGNNAALVNYDOLVTRVVTHEMAHANNTAN 149
DB 563 TGLTNN--NNSKSPAKSKKSNFD-----NNSNLANLNDKSLKINVTNITNSETSN 615

RESULT 6
US-10-170-205E-35751
; Sequence 35751, Application US/10170205E
; GENERAL INFORMATION:
; APPLICANT: ADAMS, Mark
; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN
; FILE REFERENCE: CLO01381
; CURRENT APPLICATION NUMBER: US/10/170,205E
; CURRENT FILING DATE: 2002-06-13
; NUMBER OF SEQ ID NOS: 40312
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 35751
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-170-205E-35751

Query Match 11.8%; Score 92; DB 6; Length 299;
Best Local Similarity 30.4%; Pred. No. 3;
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

QY 6 VAAFAAI-VVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQGSANALALQSDAR 64
DB 99 VAAAPVPAVSTSSAAGVAPNAPAGSGNNSPSSSPTSS-SSSPSPSGSLAESPEAA 157
QY 65 KSETTIT---QSGYNGADVGQGDNSTIELTQNG--FRNNATIDQNAKNSDI 113
DB 158 GVSSTAPLPGGAAGPGTGVPVAVSGALRELLACRNGDVSVRKRLVDAANVNAKDM 212

RESULT 7
US-10-170-205E-741
; Sequence 741, Application US/10170205E
; GENERAL INFORMATION:
; APPLICANT: ADAMS, Mark
; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN
; FILE REFERENCE: CLO01381
; CURRENT APPLICATION NUMBER: US/10/170,205E
; CURRENT FILING DATE: 2002-06-13
; NUMBER OF SEQ ID NOS: 40312
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 741
; LENGTH: 1203
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-170-205E-741

Query Match 11.8%; Score 92; DB 6; Length 1203;
Best Local Similarity 30.4%; Pred. No. 14;
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

QY 6 VAAFAAI-VVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQGSANALALQSDAR 64

DB 99 VAAAPVPAVSTSSAAGVAPNAPAGSGNNSPSSSPTSS-SSSPSPSGSLAESPEAA 157
QY 65 KSETTIT---QSGYNGADVGQGDNSTIELTQNG--FRNNATIDQNAKNSDI 113
DB 158 GVSSTAPLPGGAAGPGTGVPVAVSGALRELLACRNGDVSVRKRLVDAANVNAKDM 212

RESULT 8
PCT-US04-02338-49
; Sequence 49, Application PC/TUS0402338
; GENERAL INFORMATION:
; APPLICANT: EXELIXIS, INC.
; TITLE OF INVENTION: MAPCAXS AS MODIFIERS OF THE APC AND AXIN PATHWAYS AND METHODS OF
; FILE REFERENCE: EX04-003C-PC
; CURRENT APPLICATION NUMBER: PCT/US04/02338
; CURRENT FILING DATE: 2004-01-28
; PRIOR APPLICATION NUMBER: US60/443,484
; PRIOR FILING DATE: 2003-01-29
; PRIOR APPLICATION NUMBER: US60/447,358
; PRIOR FILING DATE: 2003-02-11
; PRIOR APPLICATION NUMBER: US60/461,789
; PRIOR FILING DATE: 2003-04-10
; PRIOR APPLICATION NUMBER: US60/470,684
; PRIOR FILING DATE: 2003-05-14
; PRIOR APPLICATION NUMBER: US60/479,650
; PRIOR FILING DATE: 2003-06-19
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 49
; LENGTH: 1327
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US04-02338-49

Query Match 11.8%; Score 92; DB 1; Length 1327;
Best Local Similarity 30.4%; Pred. No. 16;
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

QY 6 VAAFAAI-VVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQGSANALALQSDAR 64
DB 99 VAAAPVPAVSTSSAAGVAPNAPAGSGNNSPSSSPTSS-SSSPSPSGSLAESPEAA 157
QY 65 KSETTIT---QSGYNGADVGQGDNSTIELTQNG--FRNNATIDQNAKNSDI 113
DB 158 GVSSTAPLPGGAAGPGTGVPVAVSGALRELLACRNGDVSVRKRLVDAANVNAKDM 212

RESULT 9
US-10-425-115-301334
; Sequence 301334, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 301334
; LENGTH: 256
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_37894C.1.pcp
US-10-425-115-301334

Query Match 11.7%; Score 91.5; DB 6; Length 256;
Best Local Similarity 26.6%; Pred. No. 2.8;

```
Matches 33; Conservative 16; Mismatches 52; Indels 23; Gaps 5;
QY 17 SALAGVVPVWGGGNGHNGSSGPDSTLSIYQGSANAALALQSDA----RKSETTITQ 72
Db 79 SSVAG-----GGGGGGGGGTNGSGSGSGSGTSTTAASGSPSSGNVANAGKAG 133
QY 73 SGYGNAGD-----VQGGADNSTIEL--TQGNFRNATIDQWNAKNSDITVGYQGG 120
Db 134 GGMGGGADGAYSGAGGAGGKGGGSGVALAPSSDGYNGGAADATGGGSG--AGGGHGG 191
QY 121 NNA 124
Db 192 GAGA 195

RESULT 10
US-10-100-683-7608
; Sequence 7608, Application US/10100683
; GENERAL INFORMATION:
; APPLICANT: Rosen, et al.
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PS900
; CURRENT APPLICATION NUMBER: US/10/100,683
; CURRENT FILING DATE: 2002-03-19
; PRIOR APPLICATION NUMBER: US 60/040,162
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: US 60/043,576
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: US 60/047,601
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: US 60/056,845
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/043,580
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: US 60/047,599
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: US 60/056,664
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/043,314
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: US 60/047,632
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: US 60/056,892
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 13468
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7608
; LENGTH: 443
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-100-683-7608

Query Match 11.7%; Score 91.5; DB 6; Length 443;
Best Local Similarity 26.0%; Pred. No. 5.1;
Matches 33; Conservative 18; Mismatches 43; Indels 33; Gaps 6;
QY 21 GVVPVQ---WGGGNGHNGSSGPDSTLSIYQ--YGSANAALALQSDARKSETTITQSGY 75
Db 185 GMPQAPAPGQGG--NGGPPNGTNTQGAAPGPGYGVRA-----SNQNEGCTNPPPSGS 237
QY 76 GNGAD---VQGGADNSTIELTQGNFRNATI-----DQWNAKNSDIT 114
Db 238 GGGSSNGGSGSGSGSGSGSGGNDNNNGSSGSGSGSGSGSGSGSGSGSGSSSTGSS 297
QY 115 VQYQYGN 121
Db 298 SGNHGG 304

RESULT 11
US-60-565-632-7906
; Sequence 7906, Application US/60565632
```

```
; GENERAL INFORMATION:
; APPLICANT: Monsanto Technology, LLC
; APPLICANT: Baum, James A
; APPLICANT: Kovalic, David K.
; APPLICANT: Larosa, Thomas J.
; APPLICANT: Lu, Maolong
; APPLICANT: Muryikwa, Tichifa R. I.
; APPLICANT: Roberts, James K.
; APPLICANT: Wu, Wei
; APPLICANT: Zhang, Bei
; TITLE OF INVENTION: Methods for Genetic Control of Insect Infestations in Plants and
; FILE REFERENCE: 38-21(53403)B
; CURRENT APPLICATION NUMBER: US/60/565,632
; CURRENT FILING DATE: 2004-04-27
; NUMBER OF SEQ ID NOS: 15449
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7906
; LENGTH: 841
; TYPE: PRT
; ORGANISM: Diabrotica virgifera
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (810)..(810)
; OTHER INFORMATION: Xaa can be any naturally occurring amino acid
US-60-565-632-7906

Query Match 11.7%; Score 91.5; DB 7; Length 841;
Best Local Similarity 32.6%; Pred. No. 11;
Matches 44; Conservative 12; Mismatches 48; Indels 31; Gaps 9;
QY 30 GNGGGG--NSSGPDSTLSIYQGSANAALA-----LQSDARKSETTITQSGYGNAD 80
Db 269 GNGGGAENNAADACTDAAQ--GSTNEAENNAADVQNDAAQENGAENGAENSGNAD 327
QY 81 VGGADN--STIELTON-----GFRNATIDQWNAKNSDITVGYQGNNAALVNYDQVTR 134
Db 328 AAQCTDNGAAANTGNADPAQGNNGAA-----AENSGNENGTAAGNNA---NPD----- 374
QY 135 VVTHEMAHNN-NATA 148
Db 375 -VNDAAQVNDNGTA 388

RESULT 12
US-60-579-062-7906
; Sequence 7906, Application US/60579062
; GENERAL INFORMATION:
; APPLICANT: Baum, James A
; APPLICANT: Kovalic, David K
; APPLICANT: Larosa, Thomas J
; APPLICANT: Lu, Maolong
; APPLICANT: Muryikwa, Tichifa R. I.
; APPLICANT: Roberts, James K
; APPLICANT: Wu, Wei
; APPLICANT: Zhang, Bei
; TITLE OF INVENTION: Methods for Genetic Control of Insect Infestations in Plants and
; FILE REFERENCE: 38-21(53403) C
; CURRENT APPLICATION NUMBER: US/60/579,062
; CURRENT FILING DATE: 2004-06-11
; NUMBER OF SEQ ID NOS: 41445
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7906
; LENGTH: 841
; TYPE: PRT
; ORGANISM: Diabrotica virgifera
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (810)..(810)
; OTHER INFORMATION: Xaa can be any naturally occurring amino acid
US-60-579-062-7906
```


[illegible]

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:39:53 ; Search time 9.4 Seconds

(without alignments)
1545.204 Million cell updates/sec

Title: US-09-543-407-12

Perfect score: 779

Sequence: 1 MKLLKVAFAAIVVSGSALA.....VTRVVTTHMAHANNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_78.*

1: Piri.*

2: Piri2.*

3: Piri3.*

4: Piri4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	695	89.2	151	JC6039	fimbrin protein ag
2	695	89.2	151	A10635	major curlin chain
3	520	66.8	151	S70788	curlin protein csg
4	501.5	64.4	152	D90806	curlin major subun
5	501.5	64.4	152	H85665	hypothetical prote
6	106	13.6	1748	S82136	cujB protein - ret
7	104	13.4	2174	E95965	hypothetical glyci
8	101.5	13.0	151	AH0635	nucleation compone
9	100.5	12.9	151	JC6040	fimbrin protein ag
10	98.5	12.6	1028	A56038	DNA-binding protei
11	98.5	12.6	1213	S16356	ovo protein - frui
12	96	12.3	145	AD3143	conserved hypothet
13	96	12.3	145	H98144	hypothetical prote
14	94.5	12.1	678	A70762	probable PPE prote
15	93.5	12.0	1238	AH0038	probable PPE prote
16	93	11.9	573	C86266	F3f19.21 protein -
17	91.5	11.7	151	S70787	curlin nucleator p
18	91.5	11.7	151	C90806	minor curlin subun
19	91.5	11.7	151	C85665	curlin minor chain
20	91.5	11.7	256	T03371	glycine-rich prote
21	91	11.7	770	T51034	related to C2H2 z1
22	91	11.7	1250	D91018	hypothetical prote
23	91	11.7	1250	F85862	hypothetical prote
24	90.5	11.6	745	B64559	outer membrane pro
25	90	11.6	409	T20847	hypothetical prote
26	90	11.6	573	S35327	protein kinase sgg
27	90	11.6	582	F70675	probable PPE prote
28	89.5	11.5	262	S00275	tail fiber protein
29	89.5	11.5	343	T05221	hypothetical prote

RESULT 1

JC6039
fimbrin protein agfA precursor - Salmonella enteritidis
C;Species: Salmonella enteritidis
C;Date: 31-Dec-1996 #sequence revision 31-Dec-1996 #text_change 08-Oct-1999
C;Accession: JC6039; PC6015; A44898
R;Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Banser, P.A.; Kay, W.W.
J. Bacteriol. 178, 662-667, 1996
A;Title: Salmonella enteritidis agfBAC operon encoding thin, aggregative fimbrinae.
A;Reference number: JC6039; MUID:96146512; PMID:8550497
A;Accession: JC6039
A;Molecule type: DNA
A;Residues: 1-151 <COL>
A;Cross-references: GB:U43280; NID:gl184712; PIDN:ARC43599.1; PID:gl184714
A;Accession: PC6015
A;Molecule type: protein
A;Residues: 21-52 <CO2>
A;Experimental source: strain 27655-3b
A;Note: the authors translated the codon ACG for residue 44 as Ile
R;Collinson, S.K.; Emody, L.; Muller, K.H.; Trust, T.J.; Kay, W.W.
J. Bacteriol. 173, 4773-4781, 1991
A;Title: Purification and characterization of thin, aggregative fimbrinae from Salmonell
A;Reference number: A44898; MUID:91310586; PMID:1677357
A;Contents: 27655
A;Accession: A44898
A;Status: preliminary
A;Molecule type: protein
A;Residues: 21-33 <CO3>
A;Note: sequence extracted from NCBI backbone (NCBIP:45936)
C;Genetics:
A;Gene: agfA
C;Function:
A;Description: major component of thin aggregative fimbrinae
A;Note: fimbrinae bind to fibronectin, plasminogen, tissue plasminogen activator
C;Keywords: fimbrina
F;1-20/Domain: signal sequence #status predicted <SIG>
F;21-151/Product: fimbrin protein agfA #status experimental <MAT>

Query Match 89.2%; Score 695; DB 2; Length 151;

Best Local Similarity 90.1%; Pred. No. 1.8e-50;

Matches 136; Conservative 4; Mismatches 11; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALAIQ 60

DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALAIQ 60

QY 61 SPARKSETTITOSGYGNGADVCGQADNSTIETQTQGFRRNATIDOWNAKNSDITVGYGG 120

DB 61 SPARKSETTITOSGYGNGADVCGQADNSTIETQTQGFRRNATIDOWNAKNSDITVGYGG 120

QY 121 NNAALVNVYDQLVTRVVTTHMAHANNATANQY 151

DB 121 NNAALVNVYDQLVTRVVTTHMAHANNATANQY 151

Db 121 NNAALVNQTSADSSVMVRQVGFNNATANQY 151

RESULT 2

AI0635

major curlin chain precursor [imported] - Salmonella enterica subsp. enterica serovar Typhimurium

C:Species: Salmonella enterica subsp. enterica serovar Typhi

A:Note: this species has also been called Salmonella typhi

C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002

C:Accession: AI0635

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, C.; Parkhill, J.; Cronin, A.; Davies, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, M.; Mouton, R.; O'Garra, P.

A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhimurium

A:Reference number: AB0502; MUID:21534947; PMID:11677608

A:Accession: AI0635

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-151 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD08268.1; PID:g16502315; GSPDB:GN00176

C:Genetics:

A:Gene: STY1181

Query Match 89.2%; Score 695; DB 2; Length 151;

Best Local Similarity 90.1%; Pred. No. 1.8e-50;

Matches 136; Conservative 4; Mismatches 11; Indels 0; Gaps 0;

Qy 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGGPDSTLSIYQYGSANAALALQ 60

Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGGPDSTLSIYQYGSANAALALQ 60

Qy 61 SDARKSETTITQSGYNGADVQGGADNSTIETQNGFRNNATIDQWNAKNSDITVQYGG 120

Db 61 SDARKSETTITQSGYNGADVQGGADNSTIETQNGFRNNATIDQWNAKNSDITVQYGG 120

Qy 121 NNAALVNYDQLVTRVVTTHMAHANNATANQY 151

Db 121 NNAALVNQTSADSSVMVRQVGFNNATANQY 151

RESULT 3

S70788

curlin protein csga precursor - Escherichia coli (strain K-12)

N:Alternate names: csga protein; major curlin protein

C:Species: Escherichia coli

C:Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 01-Mar-2002

C:Accession: S70788; G64846; S31202; S34560; S34559

R:Hammar, M.; Arngqvist, A.; Bian, Z.; Olsen, A.; Normark, S.

Mol. Microbiol. 18, 661-670, 1995

A:Title: Expression of two csq operons is required for production of fibronectin- and collagen-binding proteins in Escherichia coli

A:Reference number: S70783; MUID:96414468; PMID:8817489

A:Accession: S70788

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-151 <HAB>

A:Cross-references: EMBL:X90754; NID:g1147558; PIDN:CAA62282.1; PID:g1147564

A:Experimental source: strain K12, substrain W3110

A:Note: The nucleotide sequence was submitted to the EMBL Data Library, August 1995

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, D.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: G64846

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-151 <BLAT>

A:Cross-references: GB:AE000205; GB:U00096; NID:g1787265; PIDN:AA074126.1; PID:g1787279;

A:Experimental source: strain K-12, substrain MG1655

R:Olsen, A.; Arngqvist, A.; Hammar, M.; Sukupolvi, S.; Normark, S.

Mol. Microbiol. 7, 523-536, 1993

A:Title: The RpoS sigma factor relieves H-NS-mediated transcriptional repression of csqA

A:Reference number: S31202; MUID:93211294; PMID:8459772

A:Accession: S31202

A:Molecule type: DNA

A:Residues: 1-6,'V', 8-151 <OLS1>

A:Cross-references: EMBL:L04979

A:Accession: S34560

A:Molecule type: protein

A:Residues: 21-42;44-50 <OLS2>

R:Olsen, A.N.; Arngqvist, A.M.

submitted to the EMBL Data Library, October 1992

A:Reference number: S34559

A:Accession: S34559

A:Molecule type: DNA

A:Residues: 1-133,'RQRSGWLW' <OLS3>

A:Cross-references: EMBL:L04979; NID:g290424; PIDN:AAA23616.1; PID:g290425

A:Experimental source: strain K-12, substrain W3110

C:Genetics:

A:Gene: csGA

A:Map position: 23.15

C:Function:

A:Description: major component of wild-type curli; interaction between csqA and csqB transcripts; curli are thin, coiled fibers expressed on the surface of Escherichia coli that and H-kininogen; in the absence of csqA, csqB can self-assemble into polymers

F:1-20/Domain: signal sequence #status predicted <SIG>

F:21-151/Product: curlin #status experimental <MAT>

Query Match 66.8%; Score 520; DB 2; Length 151;

Best Local Similarity 66.9%; Pred. No. 4.7e-36;

Matches 101; Conservative 21; Mismatches 29; Indels 0; Gaps 0;

Qy 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGGPDSTLSIYQYGSANAALALQ 60

Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGGPDSTLSIYQYGSANAALALQ 60

Qy 61 SDARKSETTITQSGYNGADVQGGADNSTIETQNGFRNNATIDQWNAKNSDITVQYGG 120

Db 61 TDARNSDLTITQGGGNGADVQGGSDSDSIDITQGFNGSATLDQWNGKNSDITVQYGG 120

Qy 121 NNAALVNYDQLVTRVVTTHMAHANNATANQY 151

Db 121 GNGAAVDQTSASNVVTVQVGFNNATANQY 151

RESULT 4

D90806

curlin major subunit csqA [imported] - Escherichia coli (strain O157:H7, substrain RIMD)

C:Species: Escherichia coli

C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001

C:Accession: D90806

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.

gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic islands

A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: D90806

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-152 <HAY>

A:Cross-references: GB:BA000007; PIDN:BA034843.1; PID:g13360880; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RIMD 050952

C:Genetics:

A:Gene: ECs1420

Query Match 64.4%; Score 501.5; DB 2; Length 152;

Best Local Similarity 65.8%; Pred. No. 1.6e-34;

Matches 100; Conservative 21; Mismatches 30; Indels 1; Gaps 1;

Qy 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGGPDSTLSIYQYGSANAALAL 59

Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGGPDSTLSIYQYGSANAALAL 60

Qy 60 QSDARKSETTITQSGYNGADVQGGADNSTIETQNGFRNNATIDQWNAKNSDITVQYGG 119

```
Db 61 QADAARNSLTIHQHGGNGADVQGGSDSSIDLQRFNGSATLDQWNGKDSHMTVKQFG 120
QY 120 GNNAAVNVYDQLVTRVTHMAHANNATANOY 151
Db 121 GNGAAVDQASNSTVNTVQVFGNNATAHQY 152

RESULT 5
H85665
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: H85665
R:Perina, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
  Miller, L.; Grobeck, B.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
  Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: H85665
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-152 <STO>
A:Cross-references: GB:AB005174; NID:gl2514574; PIDN:AAAG55788.1; GSPDB:GN00145; UWGP:216
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: csfA

Query Match 64.4%; Score 501.5; DB 2; Length 152;
Best Local Similarity 65.8%; Pred. No. 1.6e-34;
Matches 100; Conservative 21; Mismatches 30; Indels 1; Gaps 1;

QY 1 MKLLKVAFAALVVGSGALAGVPOW-CGGGNHNGGNSGPDSTLSIYQVGSANAALAL 59
Db 1 MKLLKVAFAALVVGSGALAGVPOYGGGGNGHGGGNSGPNSELNIYQVGGNSALAL 60

QY 60 QSDARKSETTITQSGYNGADVGGQADNSTIELQNGFRNNATIDQWNAKNSDITVQYG 119
Db 61 QADARNSLTITQHGGNGADVQGGSDSSIDLQRFNGSATLDQWNGKDSHMTVKQFG 120

QY 120 GNNAAVNVYDQLVTRVTHMAHANNATANOY 151
Db 121 GNGAAVDQASNSTVNTVQVFGNNATAHQY 152

RESULT 6
cnjB protein - Tetrahymena thermophila
C:Species: Tetrahymena thermophila
C:Date: 19-Mar-1997 #sequence_revision 18-Jul-1997 #text_change 07-Dec-1999
C:Accession: S42136; S42135; S03650
R:Taylor, F.M.; Martindale, D.W.
  submitted to the EMBL Data Library, October 1992
A:Reference number: S42136
A:Accession: S42136
A:Molecule type: DNA
A:Residues: 1-1748 <TAY>
A:Cross-references: EMBL:L03710; NID:gl61751; PID:gl61752
R:Taylor, F.M.; Martindale, D.W.
  Nucleic Acids Res. 21, 4610-4614, 1993
A:Title: Retroviral-type zinc fingers and glycine-rich repeats in a protein encoded by c
  submitted to the EMBL Data Library, October 1992
A:Reference number: S42135; MUID:94051569; PMID:8233798
A:Accession: S42135
A:Molecule type: DNA
A:Residues: 1154-1174; 1179-1198; 1233-1252; 1285-1293; 1297-1309; 1316-1326; 1331-1341; 1343-1
R:Martindale, D.W.; Taylor, F.M.
  Nucleic Acids Res. 16, 2189-2201, 1988
A:Title: Multiple introns in a conjugation-specific gene from Tetrahymena thermophila.
A:Reference number: S03650; MUID:88189811; PMID:3357771
A:Accession: S03650
A:Molecule type: DNA
A:Residues: 236-250, 'I', 252-255, 'N', 257-773 <MAR>
```

```
A:Cross-references: EMBL:X06462
C:Genetics:
A:Gene: cnjB
A:Genetic code: SGCS
A:Introns: 85/3; 136/1; 157/3; 201/2; 290/2; 327/3; 499/1; 573/2; 607/3; 708/3; 777/3;
C:Keywords: zinc finger
F:1164-1450/Region: glycine-rich
F:1451-1464/Region: zinc finger CCHC motif
F:1478-1491/Region: zinc finger CCHC motif
F:1501-1514/Region: zinc finger CCHC motif
F:1530-1543/Region: zinc finger CCHC motif
F:1555-1568/Region: zinc finger CCHC motif
F:1579-1592/Region: zinc finger CCHC motif
F:1602-1615/Region: zinc finger CCHC motif
F:1626-1748/Region: glycine-rich

Query Match 13.6%; Score 106; DB 2; Length 1748;
Best Local Similarity 32.1%; Pred. No. 1;
Matches 35; Conservative 13; Mismatches 33; Indels 28; Gaps 5;

QY 25 QWGGGNHNGG---GNSSGPDSTLSIYQVGSANAALALQSDARKSETTIT---QSGYGN 77
Db 1640 QFGGGNSNGSGSWGTSSGSDWN-----CQSNVQESTTSSGGWGGSGSGN 1685

QY 78 GADVGGQADNSTIELQNGFRNNATIDQWNAKNSDITVGO--VGGNNA 124
Db 1686 QTGGGWSNDN-----QQQENNTGGGGWSSNSQTNNESSWGSNNQA 1729

RESULT 7
E95965
C:Species: Sinorhizobium meliloti
C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C:Accession: E95965
R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Herna
  Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A:Title: The complete sequence of the 1.683-kb pSymb megaplasmid from the N2-fixing end
  A:Reference number: A95842; MUID:21396508; PMID:11481431
A:Accession: E95965
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2174 <KUR>
A:Cross-references: GB:AL591985; PIDN:CAC49389.1; PID:gl5140875; GSPDB:GN00167
A:Experimental source: strain 1021, megaplasmid pSymb
R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
  Peita, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.
  L.; Hyman, R.W.; Jones, T.
  Science 283, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure
  heault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
  A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
  A:Reference number: A96039; MUID:21368234; PMID:11474104
A:Contents: annotation
C:Genetics:
A:Gene: Smb21548
A:Genome: plasmid

Query Match 13.4%; Score 104; DB 2; Length 2174;
Best Local Similarity 27.1%; Pred. No. 1.9;
Matches 35; Conservative 16; Mismatches 44; Indels 34; Gaps 6;

QY 11 AIVVSGSALAGVVPQ--WGGGGNHGGGNSGPDSTLSIYQVGSANAALALQSDARKSET 68
Db 693 AIATAGAGAVGILAQISIGGGGN--GGNATGGDAGFGSFIQGGGGG----- 737

QY 69 TITQSGYNGADVQ-----QGADNSTI--ELTQNGFRNNATIDQWNAK---NSDITV 115
Db 738 ----GGYANTANVGKGLTLTTQGSAAAGIVAQVGGGGGCTGTASSYSAGIGFTASVAV 793

QY 116 GQVGGNNA 124
Db 794 GGTGGNGA 802
```

RESULT 8
AH0635
nucleation component of curlin monomers [imported] - Salmonella enterica subsp. enterica
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Notes: this species has also been called Salmonella typhi
C>Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AH0635
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AH0635
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-151 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD08267.1; PID:gl6502314; GSPDB:GN00176
C:Genetics:
A:Gene: STY1180

Query Match 13.0%; Score 101.5; DB 2; Length 151;
Best Local Similarity 30.2%; Pred. No. 0.15;
Matches 29; Conservative 17; Mismatches 49; Indels 1; Gaps 1;
QY 47 IVYGSANAALQSDARKSETTITQSGYNGADYCGQADNSTIELTQNGFNATIDOW 106
DB 49 IGQVGTDSNA-RVREGSKLLSVISQEGNNRKAQVDQAGNYFAYIEQTGNANDASISQS 107
QY 107 NAKNSDITQYQGNNAALVNDQLVTRVVTHEMAH 142
DB 108 AYGNRAITQKSGNKANITCYGTQKTAVVQKQSH 143

RESULT 9
JC6040
fimbria protein agfB precursor - Salmonella enteritidis
C:Species: Salmonella enteritidis
C>Date: 31-Dec-1996 #sequence_revision 31-Dec-1996 #text_change 08-Oct-1999
C:Accession: JC6040
R:Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Baner, P.A.; Kay, W.W.
J. Bacteriol. 178, 662-667, 1996
A:Title: Salmonella enteritidis agfBAC operon encoding thin, aggregative fimbriae.
A:Reference number: JC6039; MUID:96146512; PMID:8550497
A:Accession: JC6040
A:Molecule type: DNA
A:Residues: 1-151 <COL>
A:Cross-references: GB:U43280; NID:gl1184712; PIDN:AAC43598.1; PID:gl1184713
A:Experimental source: strain 276755-3b
C:Genetics:
A:Gene: agfB
C:Function:
A:Description: minor component of thin aggregative fimbriae
A:Note: fimbriae bind to fibronectin, plasminogen, tissue plasminogen activator
C:Keywords: fimbria
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-151/Product: fimbria protein agfB #status predicted <MAT>

Query Match 12.9%; Score 100.5; DB 2; Length 151;
Best Local Similarity 30.2%; Pred. No. 0.18;
Matches 29; Conservative 17; Mismatches 49; Indels 1; Gaps 1;
QY 47 IVYGSANAALQSDARKSETTITQSGYNGADYCGQADNSTIELTQNGFNATIDOW 106
DB 49 IGQVGTDSNA-RVREGSKLLSVISQEGNNRKAQVDQAGNYFAYIEQTGNANDASISQS 107
QY 107 NAKNSDITQYQGNNAALVNDQLVTRVVTHEMAH 142
DB 108 AYGNRAITQKSGNKANITCYGTQKTAVVQKQSH 143

RESULT 10
A56038
DNA-binding protein ovo - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C>Date: 01-Dec-1995 #sequence_revision 01-Dec-1995 #text_change 21-Jul-2000
C:Accession: A56038
R:Garfinkel, M.D.; Wang, J.; Liang, Y.; Mahowald, A.P.
Mol. Cell Biol. 14, 6809-6818, 1994
A:Title: Multiple products from the shavenbaby-ovo gene region of Drosophila melanogaster
A:Reference number: A56038; MUID:95021209; PMID:7935398
A:Accession: A56038
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1028 <GAR>
A:Cross-references: GB:U11383; NID:G520526; PIDN:AAB60216.1; PID:G520527
C:Genetics:
A:Gene: ovo
A:Cross-references: FlyBase:FBgn0003028

Query Match 12.6%; Score 98.5; DB 2; Length 1028;
Best Local Similarity 24.5%; Pred. No. 2.3;
Matches 46; Conservative 15; Mismatches 62; Indels 65; Gaps 8;
QY 3 LLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQGSANAALALQSD 62
DB 59 LQNAAAAYIMSAGS-----GGCTCGGGGASGPGGSPANSNGGGGG----- 104
QY 63 ARKSETTITQSGYNGADYCGQADNSTIELTQNGFNATIDOWNAKNSDI----- 113
DB 105 -----CGNGYINCGVG-GPNN--LDGNLLNFASVSNYNSKGFNNHHHHQH 152
QY 114 -----TVQG-----YCGNNAAL-----VNYDQ-----LVTRVVTHEMAH 143
DB 153 NNNNNNGGQTSMMVHPFYGGNPAYGIILKQEPDIEYDEAKTIDGTFAQNIQTMGSS 212
QY 144 NNATANQY 151
DB 213 GQFNASAY 220

RESULT 11
S16356
ovo protein - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C>Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 16-Feb-1997
C:Accession: S16356
R:Mevel-Ninio, M.; Terracol, R.; Kafatos, F.C.
EMBO J. 10, 2259-2265, 1991
A:Title: The ovo gene of Drosophila encodes a zinc finger protein required for female g
A:Reference number: S16356; MUID:91293102; PMID:1712294
A:Accession: S16356
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1213 <MEV>
A:Cross-references: EMBL:X59772
C:Genetics:
A:Gene: FlyBase:ovo
A:Cross-references: FlyBase:FBgn0003028
A:Introns: 931/3; 1152/3

Query Match 12.6%; Score 98.5; DB 2; Length 1213;
Best Local Similarity 24.5%; Pred. No. 2.8;
Matches 46; Conservative 15; Mismatches 62; Indels 65; Gaps 8;
QY 3 LLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQGSANAALALQSD 62
DB 422 LQNAAAAYIMSAGS-----GGCTCGGGGASGPGGSPANSNGGGGG----- 467
QY 63 ARKSETTITQSGYNGADYCGQADNSTIELTQNGFNATIDOWNAKNSDI----- 113
DB 468 -----CGNGYINCGVG-GPNN--LDGNLLNFASVSNYNSKGFNNHHHHQH 515

```
QY 114 -----TVGQ-----YGGNAAAL-----VNYDQ-----LVTRVVTHEMAHA 143
Db 516 NNNNNNGGQSMWGHPPYGNPSPAYGILLKDEPDIEYDEAKIDIGTFAQNIIOATWGSS 575
QY 144 NNATANQY 151
Db 576 GQFNASAY 583

RESULT 12
AD3143
conserved hypothetical protein Atu4768 [imported] - Agrobacterium tumefaciens (strain C58)
C:Species: Agrobacterium tumefaciens
C>Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C:Accession: AD3143
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.;
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClellan,
Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AD3143
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-145 <KUR>
A:Cross-references: GB:AR008689; PIDN:AAL45562.1; PID:g17743277; GSPDB:GN00187
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atu4768
A:Map position: linear chromosome

Query Match 12.3%; Score 96; DB 2; Length 145;
Best Local Similarity 27.1%; Pred. No. 0.41;
Matches 38; Conservative 23; Mismatches 59; Indels 20; Gaps 5;

QY 3 LLKVAFAAIVVSGSALAGVVP-----OWG-----GGNHNGGN-----SSGPDST 44
Db 1 MIRKSFASALVALVGLSAAAPAMANDVRIEYGYWSNSAGGAQEGYGNRIYQNGYNR 60
QY 45 LSIYQYSANAALQSDARKSETTITQSGYGNAGADVQGGADNSTIELTQNGFRNNATID 104
Db 61 IVGHQYGRHNLS-AVGQEGHDNYGTTQNGRNVAGIQFGSNHTTILTDGNGNIAAGV 119
QY 105 QWNAKNSDITVGYGGNNAA 124
Db 120 Q-VGRGCSANVSQGGNDNVA 138

RESULT 13
H98144
Hypothetical protein AGR_L_228 [imported] - Agrobacterium tumefaciens (strain C58, Cerec)
C:Species: Agrobacterium tumefaciens
C>Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
C:Accession: H98144
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Accession: H98144
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-145 <KUR>
A:Cross-references: GB:AE007870; PIDN:AAK88682.1; PID:g15158413; GSPDB:GN00170
C:Genetics:
A:Gene: AGR_L_228
A:Map position: linear chromosome

Query Match 12.3%; Score 96; DB 2; Length 145;
Best Local Similarity 27.1%; Pred. No. 0.41;
Matches 38; Conservative 23; Mismatches 59; Indels 20; Gaps 5;
```

```
QY 3 LLKVAFAAIVVSGSALAGVVP-----OWG-----GGNHNGGN-----SSGPDST 44
Db 1 MIRKSFASALVALVGLSAAAPAMANDVRIEYGYWSNSAGGAQEGYGNRIYQNGYNR 60
QY 45 LSIYQYSANAALQSDARKSETTITQSGYGNAGADVQGGADNSTIELTQNGFRNNATID 104
Db 61 IVGHQYGRHNLS-AVGQEGHDNYGTTQNGRNVAGIQFGSNHTTILTDGNGNIAAGV 119
QY 105 QWNAKNSDITVGYGGNNAA 124
Db 120 Q-VGRGCSANVSQGGNDNVA 138

RESULT 14
A70762
probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 15-Sep-2003
C:Accession: A70762
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon,
Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Scars, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: A70762
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-678 <COL>
A:Cross-references: GB:Z74020; GB:AL123456; NID:g3261584; PIDN:CAA98335.1; PID:e1300075
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: PPE

Query Match 12.1%; Score 94.5; DB 2; Length 678;
Best Local Similarity 25.7%; Pred. No. 3.1;
Matches 39; Conservative 24; Mismatches 48; Indels 41; Gaps 8;

QY 15 SGSALAGVVPQGGGNGHNGNSGPDSTLSIYQYSANAALQ--SDARKSETTITQ 72
Db 384 SSGSGLG----FGNSGNGIGFFNSG--NNNIGMNSGNGVGALSVEFGSAAERS----- 432
QY 73 SGYGNAGADVQGGADNS-----TIELTQNGFRNNATIDQ--WNAKNSDITVGYGGNN 122
Db 433 SGFGSGELSTGIGNSGQLSTGWFNSATTSTGWFNSGTTTTCGFWNSGTTTTCGNSGNN- 491
QY 123 AALVNYDQIVTRVVTHEM-----AHANNATAN 149
Db 492 -----LVTSGLFNSGHTNTGSPN 511

RESULT 15
AH0038
probable exported protein YPO0309 [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C>Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
C:Accession: AH0038
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tibball, R.W.; Holden, M.T.G.; Prentice, M.B.
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AH0038
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1238 <KUR>
A:Cross-references: GB:AL590842; PIDN:CAC89171.1; PID:g15978410; GSPDB:GN00175
C:Genetics:
A:Gene: YPO0309
```

```

Query Match      12.0%; Score 93.5; DB 2; Length 1238;
Best Local Similarity 25.1%; Pred. No. 7.3;
Matches 45; Conservative 19; Mismatches 48; Indels 67; Gaps 9;

QY 7 AAFALVSGSALAGVVPWGSGG---NHNGGN-----SSG-----P 41
Db 409 SAFSAITATGHLTA-----EWGQGLQTHSSSLGDATLLHFNDITAMSSGISLINEANQG 464
QY 42 DSTLSIVQYGSANAA-----LALQSDARKSETTITQSGYGN 77
Db 465 TSTADITVIGQINVSHGEGITLNALTTDGRTLVNVVDVNNIASEYDAIRLYNNYNDNYAT 524
QY 78 GADVGGQADN--STIEL-----TQNGFRNNATIDQWNAKNSDITVG--QYGGNNAAL 125
Db 525 GVDDGTGADNGTSTDLITRGALVSGQGYGINI---ETNTADTYVTVGGLVHGGNGTAI 580

```

Search completed: August 2, 2004, 14:56:21
Job time : 10.4 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run On: August 2, 2004, 14:36:12 ; Search time 5.3 Seconds
(without alignments)

1483.508 Million cell updates/sec

Title: US-09-543-407-12

Perfect score: 779

Sequence: 1 MKLLKVAFAAIVVSGSALA.....VTRVVTHEMAFANNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	695	89.2	151	1 CSGA_SALTY	P5225 salmonella
2	520	66.8	151	1 CSGA_ECOLI	P28307 escherichia
3	501.5	64.4	152	1 CSGA_ECO57	Q91u24 escherichia
4	101.5	13.0	151	1 CSGB_SALTY	Q827m3 salmonella
5	100.5	12.9	151	1 CSGB_SALTY	P5225 salmonella
6	98.5	12.6	1028	1 OVO_DROME	P51521 drosophila
7	97	12.5	1656	1 OMPB_RICJA	O06653 r outer mem
8	94.5	12.1	678	1 YF48_MYCTU	Q10778 mycobacteri
9	92	11.8	1327	1 TNK1_HUMAN	O85271 homo sapien
10	91.5	11.7	151	1 CSGB_ECOLI	P39828 escherichia
11	89.5	11.5	262	1 VG38_EPT2	P07875 bacterioph
12	89	11.4	347	1 MGA2_PLAF2	Q03646 plasmodium
13	88	11.3	491	1 YK98_MYCTU	Q10707 mycobacteri
14	87	11.2	331	1 OMB2_NEIMB	P30688 neisseria m
15	86	11.0	1093	1 PER_DROWI	Q03297 drosophila
16	85	10.9	1250	1 YFAL_ECOLI	P45508 escherichia
17	84.5	10.8	3590	1 FHAB_BORPE	P12255 bordetella
18	84	10.8	566	1 AMY_STRGR	P30270 streptomyce
19	84	10.8	566	1 AMY_STRLM	P09794 streptomyce
20	83.5	10.7	1067	1 CSG_DROME	P18431 drosophila
21	83	10.7	590	1 GP63_LEIDO	P23223 leishmania
22	83	10.7	599	1 GP63_LEIGH	P15706 leishmania
23	83	10.7	602	1 GP63_LEIMA	P08148 leishmania
24	83	10.7	646	1 GP63_LEIME	P43150 leishmania
25	82.5	10.6	369	1 PST3_MYCAV	Q9Kk89 mycobacteri
26	82.5	10.6	794	1 YC84_MYCPN	P75493 mycoplasma
27	82	10.5	1567	1 ICEN_XANCT	P18127 xanthomonas
28	81.5	10.5	392	1 HME1_HUMAN	Q05925 homo sapien
29	81.5	10.5	730	1 GLN3_YEAST	P18494 saccharomyc
30	81	10.4	165	1 GRP1_ORYSA	P25074 oryza sativ
31	81	10.4	385	1 PER_DRONE	P91686 drosophila
32	81	10.4	485	1 YB47_MYCPN	P51339 mycoplasma
33	81	10.4	959	1 N100_YEAST	Q02629 saccharomyc

34 81 10.4 1460 1 PMPC_CHLMU
35 80.5 10.3 487 1 Y442_MYCTU
36 80.5 10.3 2003 1 YDBA_ECOLI
37 80 10.3 362 1 P35_MYCPE
38 80 10.3 594 1 SUH_DROME
39 80 10.3 955 1 FRU_DROME
40 80 10.3 1778 1 N189_SCHPO
41 79.5 10.2 485 1 Y136_TREPA
42 79.5 10.2 760 1 YB1L_ECOLI
43 79 10.1 572 1 FLAA_CAMCO
44 79 10.1 933 1 NPA3_HUMAN
45 79 10.1 1113 1 N116_YEAST

ALIGNMENTS

RESULT 1
CSGA_SALTY STANDARD; PRT; 151 AA.
AC P5225;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Major curlin subunit precursor (Fimbrin SEF17).
GN CSGA OR AGFA OR STM144 OR STY1181 OR T1776.
OS Salmonella typhimurium,
OS Salmonella typhi, and
OS Salmonella enteritidis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602, 601, 592;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=SR-11;
RX MEDLINE=98117058; PubMed=9457880;
RA Romling U., Bian Z., Hammar M., Sierralta W.D., Normark S.;
RT "Curli fibers are highly conserved between Salmonella typhimurium and
Escherichia coli with respect to operon structure and regulation.";
RL J. Bacteriol. 180:722-731(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courteney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
LT2.";
RL Nature 413:852-856(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies K.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrett B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
RN [4]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;
RA Deng W., Liu S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kocoyanni V., Schwartz D.C., Blattner F.R.;

Q9piy1 chlamydia m
P42611 mycobacteri
P33666 escherichia
Q50367 mycoplasma
P28159 drosophila
Q81n81 drosophila
Q9utk4 schizosacch
O83172 treponema p
P75780 escherichia
P27053 campylobact
O8ixf0 homo sapien
Q02630 saccharomyc

RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
 RL and CT18.";
 RL J. Bacteriol. 185:2330-2337(2003).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S. enteritidis; STRAIN=27655-3B;
 RX MEDLINE=96146512; PubMed=8550497;
 RA Collinson S.K., Clouthier S.C., Doran J.L., Baner P.A., Kay W.W.;
 RT "Salmonella enteritidis agfBAC operon encoding thin, aggregative
 RT fimbriae";
 RL J. Bacteriol. 178:662-667(1996).
 RN [6]
 RP SEQUENCE OF 21-151 FROM N.A.
 RC SPECIES=S. enteritidis; STRAIN=27655-3B;
 RX MEDLINE=91310586; PubMed=1677357;
 RA Collinson S.K., Emeody L., Mueller K.-M., Trust T.J., Kay W.W.;
 RT "Purification and characterization of thin, aggregative fimbriae from
 RT Salmonella enteritidis";
 RL J. Bacteriol. 173:4773-4781(1991).
 CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 CC FIBRONECTIN.
 CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AJ002301; CAA05317.1; -;
 DR EMBL; A3008749; AAL20074.1; -;
 DR EMBL; AL627269; CAD08268.1; -;
 DR EMBL; A3016840; AAO69399.1; -;
 DR EMBL; U43280; AAC43599.1; -;
 DR FIC; JC6039; JC6039.
 DR StyGene; SG10608; csGA.
 KW Fimbria; Signal; Complete proteome.
 FT SIGNAL 1 20
 FT CHAIN 21 151 MAJOR CURLIN SUBUNIT.
 FT CHAIN 134 151 SVMVQVGFNNATANYQ -> DSYTQVAS (IN
 FT CONFLICT 134 151 REF. 6).
 FT SEQUENCE 151 AA; 15305 MW; B7DAC0D16B621359 CRC64;
 SQ
 Query Match 89.2%; Score 695; DB 1; Length 151;
 Best Local Similarity 90.1%; Pred. No. 4.6e-51;
 Matches 136; Conservative 4; Mismatches 11; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVSGSALAGVVPQWGGGNGGNSGGPSTLSIYQYGSANAALAQ 60
 DB 1 MKLLKVAFAAIVSGSALAGVVPQWGGGNGGNSGGPSTLSIYQYGSANAALAQ 60
 QY 61 SDARKSETTITGSGYNGADVCGGADNSTIETQNGFRNNATIDQWNAKNSDITVQYGG 120
 DB 61 SDARKSETTITGSGYNGADVCGGADNSTIETQNGFRNNATIDQWNAKNSDITVQYGG 120
 QY 121 NNAALVNYDQVTRVVTTHMAHNNATANYQ 151
 DB 121 NNAALVNYDQVTRVVTTHMAHNNATANYQ 151
 DB 121 NNAALVNYDQVTRVVTTHMAHNNATANYQ 151

RESULT 2
 CSGA_ECOLI
 ID _CSGA_ECOLI STANDARD; PRT; 151 AA.
 AC P28307;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Major curlin subunit precursor.
 GN CSGA OR B1042.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / W3110;
 RX MEDLINE=93211294; PubMed=8459772;
 RA Olsen A., Arngvist A.;
 RT "The RpoS sigma factor relieves H-NS-mediated transcriptional
 RT repression of csgA, the subunit gene of fibronectin-binding curli in
 RT Escherichia coli.";
 RL Mol. Microbiol. 7:523-536(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MC4100;
 RX MEDLINE=96414468; PubMed=8817489;
 RA Hammar M., Arngvist A., Sian Z., Olsen A., Normark S.;
 RT "Expression of two csg operons is required for production of
 RT fibronectin- and congo red-binding curli polymers in Escherichia coli
 RT K-12.";
 RL Mol. Microbiol. 18:661-670(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97061202; PubMed=8905232;
 RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
 RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horiuchi T.;
 RT "A 7.18-kb DNA sequence of the Escherichia coli K-12 genome
 RT corresponding to the 12.7-28.0 min region on the linkage map.";
 RL DNA Res. 3:137-155(1996).
 RN [5]
 RP SEQUENCE OF 21-40.
 RC STRAIN=K12 / YNEL;
 RX MEDLINE=93023873; PubMed=1357528;
 RA Arngvist A., Olsen A., Pfeifer J., Russell D.G., Normark S.;
 RT "The Crl protein activates cryptic genes for curli formation and
 RT fibronectin binding in Escherichia coli HB101.";
 RL Mol. Microbiol. 6:2443-2452(1992).
 RN [6]
 RP SEQUENCE OF 21-31.
 RC MEDLINE=91310586; PubMed=1677357;
 RA Collinson S.K., Emeody L., Trust T.J., Kay W.W.;
 RT "Purification and characterization of thin, aggregative fimbriae from
 RT Salmonella enteritidis";
 RL J. Bacteriol. 173:4773-4781(1991).
 CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 CC FIBRONECTIN.
 CC

```
CC CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outpost on
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to licenses@isb-sib.ch).
CC -----
CC EMBL; L04979; AAA23616.1; -.
DR DR EMBL; X90754; CAA62282.1; -.
DR DR EMBL; AE000205; AAC74126.1; -.
DR DR EMBL; D90741; BAA35832.1; -.
DR DR EMBL; D90742; BAA35840.1; -.
DR DR PIR; S70788; S70788.
DR DR ECoGene; EG11489; csGA.
KW KW Fimbria; Signal; Complete proteome.
FT SIGNAL 1 20
FT CHAIN 21 151 MAJOR CURLIN SUBUNIT.
FT CONFLICT 7 A->E (IN REF. 1).
FT SEQUENCE 151 AA; 15049 MW; C003470D208D395F CRC64;
SQ
Query Match 66.8%; Score 520; DB 1; Length 151;
Best Local Similarity 66.9%; Pred.No.1.5e-36;
Matches 101; Conservative 21; Mismatches 29; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIIVVSGNALGVVPQGCGGNHNGGNSGGPDSTLSIYOVSANAALAIQ 60
DB 1 MKLLKVAFAAIIVVSGNALGVVPQYGCGGHGCGNNGSPNELNIYYOYGGSSALLAIQ 60
QY 61 SPARKSETTTIQSYGVNGADVGQAGDNSTIELTQGFRRNNATIDQWNAKSNDITVGYGG 120
DB 61 TDAENSDLTITQHGGENGADVQGSDDSSDLTGQFGNSATLDQWNKNSEMTVKQFGG 120
QY 121 NMAALVVYDQLVTRVTTHMAMHNATANQY 151
DB 121 NGGAAVDOTASNSSVNVTQVFGNNATAHQY 151
RESULT 3
CSGA_ECO57 STANDARD; PRY; 152 AA.
AC Q93024; AC
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Major curlin subunit precursor.
OS CSga OR Z1676 OR ECSJ420.
GN Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxId=83334; [1]
RX
RN SEQUENCE FROM N.A. [2]
RP STRAIN=O157:H7 / ATCC 43895;
RC MEDLINE=21218556; PubMed=11319125;
RX Ullrich G.A., Keen J.E., Elder R.O.;
RT "Mutations in the csGD promoter associated with variations in curlI
expression in certain strains of Escherichia coli O157:H7."
RL Appl. Environ. Microbiol. 67:2367-2370(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RX Perna N.T., Plunkett G. III, Burland V., Mau B., Glaesner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimailanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
RN Nature 409:529-533(2001).
```


SEQUENCE FROM N.A.
RC TISSUE=Ovary;
RX MEDLINE=95021209; PubMed=7935398;
RA Garinkel M.D., Wang J., Liang Y., Mahowald A.P.;
RT "Multiple products from the shavenbaby-ovo gene region of Drosophila
melanogaster: relationship to genetic complexity";
RL Mol. Cell. Biol. 14:6809-6818(1994).
RN [2].

SEQUENCE FROM N.A.
RP STRAIN=Oregon-R;
RX MEDLINE=91293102; PubMed=1712294;
RA Mevel-Ninio M.T.M., Terracol R., Kafatos F.C.;
RT "The ovo gene of Drosophila encodes a zinc finger protein required
for female germ line development.";
RL EMOB J. 10:2259-2266(1991).
RN [2].

FUNCTION: REQUIRED FOR SURVIVAL AND DIFFERENTIATION OF FEMALE GERM
LINE CELLS. PLAYS A ROLE IN GERM LINE SEX DETERMINATION.
SUBCELLULAR LOCATION: Nuclear (Potential).
DEVELOPMENTAL STAGE: FIRST APPEARS IN THE GERMARIUM AND
ACCUMULATES IN NURSE CELLS DURING OÖGENESIS. STORED IN THE EGG,
BUT IS RAPIDLY LOST IN THE EMBRYOS EXCEPT FOR ITS CONTINUED
PRESENCE IN THE GERM LINE PRECURSOR POLE CELLS.
SIMILARITY: Contains 4 C2H2-type zinc fingers.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; U11383; AAB60216.1; -;
EMBL; X59772; CAB36921.1; ALT_SEQ.
PIR; A56038; A56038.
HSP; P07248; 2ADR.
TRANSFAC; T00669; -;
FLYBASE; FBgn0003028; ovo.
InterPro; IPR007087; Znf_C2H2.
Pfam; PF00096; zf-C2H2; 3.
SMART; SM00355; Znf_C2H2; 4.
PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
PROSITE; PS00157; ZINC_FINGER_C2H2_2; 3.
Zinc-finger; Metal-binding; DNA-binding; Repeat; Nuclear protein;
Transcription regulation.
DOMAIN 62 66 POLY-ALA.
DOMAIN 72 77 POLY-GLY.
DOMAIN 80 85 POLY-GLY.
DOMAIN 98 108 POLY-GLY.
DOMAIN 144 152 POLY-HIS.
DOMAIN 153 159 POLY-ASN.
DOMAIN 336 339 POLY-GLN.
DOMAIN 347 353 POLY-GLN.
DOMAIN 357 361 POLY-GLN.
DOMAIN 410 414 POLY-GLN.
DOMAIN 418 422 POLY-GLN.
DOMAIN 426 432 POLY-GLN.
DOMAIN 445 453 POLY-GLN.
DOMAIN 456 459 POLY-GLN.
DOMAIN 466 474 POLY-GLN.
DOMAIN 497 517 POLY-ALA.
DOMAIN 524 529 POLY-SER.
DOMAIN 549 558 POLY-ALA.
DOMAIN 639 651 POLY-ALA.
DOMAIN 717 725 POLY-ALA.
DOMAIN 797 802 POLY-GLN.
DOMAIN 820 823 POLY-GLN.
DOMAIN 826 832 POLY-GLN.
ZN_FING 874 896 C2H2-TYPE 1.
ZN_FING 902 924 C2H2-TYPE 2.
ZN_FING 930 953 C2H2-TYPE 3.
ZN_FING 969 992 C2H2-TYPE 4.
CONFLICT 647 A -> R (IN REF. 2).

SQ SEQUENCE 1028 AA; 110620 MW; D7068BB2BC0F6F77 CRC64;
Query Match 12.6%; Score 98.5; DB 1; Length 1028;
Best Local Similarity 24.5%; Pred No. 1.1;
Matches 46; Conservative 15; Mismatches 62; Indels 65; Gaps 8;
QY 3 LLKVAFAAIVVSGSALAGVPMQGGGNGHNGGNSGPDSTLSIYQYGSANAALALQSD 62
DB 59 LQVAAAAAYIMSAGSG-----GGGCTGNGGGGASGPGGGSANSGGGGGG----- 104
QY 63 ARKSETTITQSGYNGADYVCGQADNSTLTITGNGFRNNATIDQWNAKQSDI----- 113
DB 105 -----GGNGYIINGGVG-GFNNS-----LDGNLLNFASVSNYNESKFNHHHHHGH 152
QY 114 -----TVGQ-----YGGNNAAL-----VNYDQ-----LVTRVVTHEMAHA 143
DB 153 NNNNNNGQTSWMGHFFYGGNPSAYGIILKDEPDIEYDEAKIDGTFAQNIQTMGSS 212
QY 144 NNATANQY 151
DB 213 GQFNASAY 220

RESULT 7
CMPB_RICJA STANDARD; PRT; 1656 AA.
ID OMPB_RICJA
AC O06653;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Outer membrane protein B precursor (168 kDa surface-layer protein)
DE (Surface protein antigen) (Cell surface antigen 5) (Scas) (rOMPb)
DE (rOMP B) [Contains: 120 kDa surface-exposed protein (Surface protein
antigen) (120 kDa outer membrane protein ompb); 32 kDa beta peptide].
GN OMPB.
OS Rickettsia japonica.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=35790;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YH;
RA Uchiyama T.;
RT "Sequencing of the gene encoding the protein rOMP B of Rickettsia
japonica";
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THE 120 kDa SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (BY
CC similarity).
CC -!- FUNCTION: THE 32 kDa BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR
CC (BY similarity).
CC -!- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-
CC layer with hexagonal symmetry.
CC -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMP FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

EMBL; AB003681; BAB20138.1; -;
DR InterPro; IPR008315; Autotransport.
DR InterPro; IPR005546; Autotransporter.
DR Pfam; PF03797; Autotransporter; 1.
DR TIGRFAMs; TIGR01414; autotrans_bar1; 2.
KW Antigen; S-layer; Cell wall.
FT CHAIN 1 1338 120 kDa SURFACE-EXPOSED PROTEIN.
FT CHAIN 1339 1656 32 kDa BETA PEPTIDE.
FT DOMAIN 528 533 POLY-GLY.

SQ SEQUENCE 1656 AA; 168097 MW; 3132A69CDD5999F CRC64;
 Query Match 12.5%; Score 97; DB 1; Length 1656;
 Best Local Similarity 28.3%; Pred. No. 2.6;
 Matches 43; Conservative 19; Mismatches 52; Indels 38; Gaps 7;
 QY 6 VAFAAIVSGSALAGVWPWGCGGHHNGGSSGPDSTLSIYQYSANAALALQSDARK 65
 DB 509 VLAAGAITDGSATI-----TGDICNGGG-----GAALQSITLANDATK 547
 QY 66 SETTITQSG---YNGADVGQAGDNSTIETQNGFRNNATIDQWNAKNSDITVVG--QVG 119
 DB 548 ---TLILGGANIISANGTINFQANGTKILST--QNNIVD-----CDLAIATDQTG 596
 QY 120 GNAALVNVYDQVTRVVVTHEMAHANNATANQY 151
 DB 597 VWDASSLTAQTLTISGTIGIIGANNTILGQF 628
 RESULT 8
 YF48_MYCTU STANDARD; PRT; 678 AA.
 ID YF48_MYCTU STANDARD; PRT; 678 AA.
 AC Q10778;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Hypothetical PPE-family protein Rv1548c/MT1599.
 GN Rv1548C OR MT1599 OR MTC148.17.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekai F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RA "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence.";
 RL Nature 393:537-544(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RX MEDLINE=22206494; PubMed=12218036;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey B.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
 RA Delcher A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
 RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains.";
 RL J. Bacteriol. 184:5479-5490(2002).
 CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; Z74020; CA98335.1; -
 CC EMBL; AE007026; AAK45866.1; ALT_INIT.
 CC EMBL; A70762; A70762.
 CC TIGR; MT1599; -

DR TubercuList; Rv1548c; -
 DR InterPro; IPR000030; Microbac_PPE.
 DR InterPro; IPR002989; Mycobac_Pentapep.
 DR Pfam; PF01469; Pentapeptide_2; 11.
 DR Pfam; PF00823; PPE; 1.
 KW Hypothetical protein; Transmembrane; Complete proteome.
 FT TRANSMEM 14 34 POTENTIAL.
 FT TRANSMEM 180 200 POTENTIAL.
 FT CONFLICT 258 258 D -> G (IN REF. 2).
 SQ SEQUENCE 678 AA; 66736 MW; 209F1593D52533A2 CRC64;
 Query Match 12.1%; Score 94.5; DB 1; Length 678;
 Best Local Similarity 25.7%; Pred. No. 1.5;
 Matches 39; Conservative 24; Mismatches 48; Indels 41; Gaps 8;
 QY 15 SGSALAGVFPWGCGGHHNGGSSGPDSTLSIYQYSANAALALQ--SDARKSETTITQ 72
 DB 384 SGSGNLG----FGNSGNGNIGFFNSG--NNNIGNSGNGVGLSVFSGSAERS----- 432
 QY 73 SGYNGADVGQAGDNS-----TIELTQNGFRNNATIDQ--WNAKNSDITVQVGGNN 122
 DB 433 SGFGNSGELSTGIGNSQQLSTGFWNSGATTSTGFWNSGTTNTGFWNSGTTNTGIGNSGNN- 491
 QY 123 AALVNVYDQVTRVVTHEM-----AHANNATAN 149
 DB 492 -----LVTGSMGLFNSGHTNTGSGFN 511
 RESULT 9
 TNK1_HUMAN STANDARD; PRT; 1327 AA.
 ID TNK1_HUMAN STANDARD; PRT; 1327 AA.
 AC O95271; O95272;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Tankyrase 1 (EC 2.4.2.30) (TANK1) (Tankyrase I) (TNKS-1) (TRF1-
 DE interacting ankyrin-related ADP-ribose polymerase).
 GN TNKS OR TNKS1 OR TIN1 OR TINF1 OR PARPL.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
 RC TISSUE=Testis;
 RX MEDLINE=99040105; PubMed=9822378;
 RA Smith S., Giriat I., Schmitt A., de Lange T.;
 RT "Tankyrase, a poly(ADP-ribose) polymerase at human telomeres.";
 RL Science 282:1484-1487(1998).
 RN [2]
 RP SUBCELLULAR LOCATION.
 RX MEDLINE=99454782; PubMed=10523501;
 RA Smith S., de Lange T.;
 RT "Cell cycle dependent localization of the telomeric PAPP, tankyrase,
 RT to nuclear pore complexes and centrosomes.";
 RL J. Cell Sci. 112:3649-3656(1999).
 RN [3]
 RP FUNCTION, AND PHOSPHORYLATION.
 RX MEDLINE=20556282; PubMed=10988299;
 RA Chi N.-W., Lodish H.F.;
 RT "Tankyrase is a Golgi-associated mitogen-activated protein kinase
 RT substrate that interacts with IRAP in GLUT4 vesicles.";
 RL J. Biol. Chem. 275:38437-38444(2000).
 RN [4]
 RP FUNCTION, AND MUTAGENESIS OF HIS-1184 AND GLU-1291.
 RX MEDLINE=21602874; PubMed=11739745;
 RA Cook B.D., Dynsek J.N., Chang W., Shostak G., Smith S.;
 RT "Role for the related poly(ADP-Ribose) polymerases tankyrase 1 and 2
 RT at human telomeres.";
 RL Mol. Cell. Biol. 22:332-342(2002).
 CC -!- FUNCTION: May regulate vesicle trafficking and modulate the
 CC subcellular distribution of SLC2A4/GLUT4-vesicles. Has PARP
 CC activity and can modify TRF1, and thereby contribute to the

CC regulation of telomere length.
CC -!- CATALYTIC ACTIVITY: NAD(+) + {ADP-D-ribose} (N)-acceptor =
CC nicotinamide + {ADP-D-ribose} (N+1)-acceptor.
CC -!- SUBUNIT: Oligomerizes and associates with TNK2. Interacts with
CC the cytoplasmic domain of INPEP/Oase in SLC2A4/GLUT4-vesicles.
CC Binds to the N-terminus of telomeric TRF1 via the ANK repeats.
CC -!- SUPRACELLULAR LOCATION: Cytoplasmic, associated with the Golgi and
CC with juxtanuclear SLC2A4/GLUT4-vesicles. A minor proportion is
CC also found at nuclear pore complexes and around the pericentriolar
CC matrix of mitotic centrosomes. During interphase, a small fraction
CC of TNKS is found in the nucleus, associated with TRF1.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=O95271-1; Sequence=Displayed;
CC Name=2;
CC IsoId=O95271-2; Sequence=VSP_004538, VSP_004539;
CC Note=No experimental confirmation available;
CC -!- TISSUE SPECIFICITY: Ubiquitous; highest levels in testis.
CC -!- PTM: Upon insulin-stimulation, phosphorylated on serine residues
CC by MARK kinases.
CC -!- PTM: ADP-ribosylated (-auto).
CC -!- SIMILARITY: Belongs to the PARP family.
CC -!- SIMILARITY: Contains 15 ANK repeats.
CC -!- SIMILARITY: Contains 1 sterile alpha motif (SAM) domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF082556; AAC79841.1; -
CC EMBL; AF082557; AAC79842.1; -
CC EMBL; AF082558; AAC79843.1; -
CC EMBL; AF082559; AAC79844.1; -
CC HSSP; Q00420; IAWC.
CC Genew; HGNC:11941; TNKS.
CC MIM; 603303; -
CC GO; GO:0000781; C:chromosome, telomeric region; IDA.
CC GO; GO:0003950; F:NAD AP-ribosyltransferase activity; IDA.
CC GO; GO:0005515; F:protein binding; IPI.
CC GO; GO:007004; P:telomerase-dependent telomere maintenance; NAS.
CC InterPro; IPR002110; ANK.
CC InterPro; IPR001660; SAM.
CC Pfam; PF00023; ank; 19.
CC Pfam; PF00536; SAM; 1.
CC PRINTS; PR01415; ANKYRIN.
CC SMART; SM00248; ANK; 17.
CC SMART; SM00454; SAM; 1.
CC PROSITE; PS0297; ANK REP REGION; 1.
CC PROSITE; PS0105; SAM DOMAIN; 1.
CC TRANSFERASE; Glycosyltransferase; NAD; Golgi stack; Telomere;
CC Nuclear protein; Repeat; ANK repeat; ADP-ribosylation;
CC Phosphorylation; Alternative splicing.
CC FT REPEAT 215 247 ANK 1.
CC FT REPEAT 248 280 ANK 2.
CC FT REPEAT 281 313 ANK 3.
CC FT REPEAT 368 400 ANK 4.
CC FT REPEAT 401 433 ANK 5.
CC FT REPEAT 434 466 ANK 6.
CC FT REPEAT 521 556 ANK 7.
CC FT REPEAT 557 589 ANK 8.
CC FT REPEAT 590 622 ANK 9.
CC FT REPEAT 683 715 ANK 10.
CC FT REPEAT 716 748 ANK 11.
CC FT REPEAT 749 781 ANK 12.
CC FT REPEAT 836 868 ANK 13.
CC FT REPEAT 869 901 ANK 14.
CC FT REPEAT 902 934 ANK 15.

FT DOMAIN 1030 1089 SAM.
FT DOMAIN 1176 1327 PARP.
FT DOMAIN 9 14 POLY-HIS.
FT DOMAIN 27 34 POLY-PRO.
FT DOMAIN 128 134 POLY-SER.
FT DOMAIN 137 145 POLY-SER.
FT VARSPPLIC 641 643 EST -> GHS (in isoform 2).
FT VARSPPLIC 644 1327 /FTid=VSP_004538.
FT VARSPPLIC 644 1327 Missing (in isoform 2).
FT MUTAGEN 1184 1184 H->A: LOSS OF ACTIVITY, WHEN ASSOCIATED
FT MUTAGEN 1291 1291 WITH A-1291.
FT MUTAGEN 1291 1291 E->A: LOSS OF ACTIVITY, WHEN ASSOCIATED
FT SEQUENCE 1327 AA; 142010 MW; E14DE985C710B957 CRC64;
Query Match 11.8%; Score 92; DB 1; Length 1327;
Best Local Similarity 30.4%; Pred. No. 5.2; Gaps 4;
Matches 35; Conservative 15; Mismatches 57; Indels 8;
QY 6 VAAFAAI-VVGSALAGVVPQWGGNGHNGSGDSTLSIYQVGSNAALALQSDAR 64
Db 99 VAAAPVPAVSTSSAAGVAPNPAGSGNNSPSSSSPTSS-SSSSPSSPSSLAESFEA 157
QY 65 KSETTIT----QSGYNGADVGOGADNSITLTQNG--FNNATIDQWNAKNSDI 113
Db 158 GVSSTAPLPGGAAGPGTGVPAVSGALRELLACRGDVSRRKLVDAANVNAKDM 212
RESULT 10
CSGB_ECOLI STANDARD; PRT; 151 AA.
AC PJ9828;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Minor curlin subunit precursor.
GN CSGB OR: B1041 OR Z1675 OR ECS1419.
OS Escherichia coli, and
OC Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 83334;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=K12 / MC4100;
RX MEDLINE=96414468; PubMed=8817489;
RA Hammar M., Arngvist A., Bian Z., Olsen A., Normark S.;
RA Blatter P.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
[3]
RN SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuko S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RT "A 718-Kb DNA sequence of the Escherichia coli K-12 genome

corresponding to the 12.7-28.0 min region on the linkage map.";
[4]
RT RC SEQUENCE FROM N.A.
RN STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G., Burland V., Mau B., Glasner J.D.,
Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoudis K.,
Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
Welch R.A., Blattner F.R.;
RA "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
RT Nature 409:529-533(2001).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RA "Complete genome sequence of enterohemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12";
RT DNA Res. 8:11-22(2001).
RN [6]
RP SEQUENCE OF 1-21 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=95157246; PubMed=7854117;
RA Arngvist A., Olsen A., Normark S.;
RT "Sigma S-dependent growth-phase induction of the csgBA promoter in
Escherichia coli can be achieved in vivo by sigma 70 in the absence
of the nucleoid-associated protein H-NS";
RT Mol. Microbiol. 13:1021-1032(1994).
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
CURLIN MONOMERS.
CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X90754; CAA62281.1; -;
DR EMBL; AB000205; AAC74125.1; -;
DR EMBL; D90741; BAA35831.1; -;
DR EMBL; AB005315; AAG55787.1; -;
DR EMBL; AP002554; BAB34842.1; -;
DR PIR; C90806; C90806;
DR PIR; G85665; G85665;
DR PIR; S70787; S70787;
DR EcoGene; EG12621; csgB.
KW Fimbria; Signal; Complete proteome.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 151 MINOR CURLIN SUBUNIT.
SQ SEQUENCE 151 AA; 15882 MW; B18D266B964014B8 CRC64;
Query Match 11.7%; Score 91.5; DB 1; Length 151;
Best Local Similarity 29.1%; Pred. No. 0.48;
Matches 32; Conservative 17; Mismatches 54; Indels 7; Gaps 4;
QY 9 FBAIVVGSAL--AGVVPWGGGNGHNGSSGPDSTLSIY-QYGSANAALALQSDARK 65
DB 33 FAVNELSKSFNQAIHII---GQAGTNNSAQLPQGGSKLLAVVQGGSSNRA-KIDQTGDY 89
QY 66 SETTITQCYGNGADVGGQADNSKTIELTQNGFRNATIDQWNAKNSDITV 115

Db 89 NLAYIDQAGSANDASISQAYGNTAMTIQKSGNKNITQYGTQKTAIVV 138
RESULT 11
VG38_BPT2
ID VG38_BPT2 STANDARD; PRT; 262 AA.
AC P07875;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 01-AUG-1990 (Rel. 15, Last annotation update)
DE Receptor recognizing protein (Protein Gp38).
GN 38.
OS Bacteriophage T2.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
OC T4-like viruses.
OX NCBI_TaxID=10664;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87283911; PubMed=3302276;
RA Riede I., Drexler K., Eschbach M.L., Henning U.;
RT "DNA sequence of genes 38 encoding a receptor-recognizing protein of
bacteriophages T2, K3 and of K3 host range mutants";
RL J. Mol. Biol. 194:31-39(1987).
CC -!- FUNCTION: VG38 is at the tip of the long tail fibers and serves as
the phage recognition site for the cellular receptor.
CC -!- MISCELLANEOUS: THIS PHAGE USE OUTER MEMBRANE PROTEINS OMPF AND TTR
AS RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR PIR; X05312; CAA28935.1; -;
DR PIR; S00275; S00275.
DR InterPro; IPR007932; Tail_fibre_GP38.
DR Pfam; PF05268; GP38; 1.
KW Fiber protein; Phage recognition.
SQ SEQUENCE 262 AA; 25801 MW; 0567366918F6C745 CRC64;
Query Match 11.5%; Score 89.5; DB 1; Length 262;
Best Local Similarity 34.1%; Pred. No. 1.3;
Matches 30; Conservative 8; Mismatches 33; Indels 17; Gaps 4;
QY 27 GGGGNGHNGGSSGPDSTLSIYQYGSANAALALQSDARKSETTITQSYGNGADVGGQAD 86
DB 175 GGGGRPFVGVGKIGSDSILS-----GSNASL---TDAGTGGTTF-QYGAGNGNGVAGGG 225
QY 87 NSTIELTQNGFRNATIDQWNAKNSDIT 114
DB 226 -----RWGKNVYTSGGAGAAVT 245
RESULT 12
MSA2_PLAF2
ID MSA2_PLAF2 STANDARD; PRT; 347 AA.
AC Q03646;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Merozoite surface antigen 2 precursor (MSA-2).
GN MSA2.
OS Plasmodium falciparum (isolate Nig32 / Nigeria).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=70150;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91156685; PubMed=2000383;
RA Smythe J.A., Coppel R.L., Day K.P., Martin R.K., Oduola A.M.J.,
RA Kemp D.J., Anders R.F.;

[illegible]

Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davis K., Devlin J., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544 (1998).
[2]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22266494; PubMed=1218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J.F., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.J., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL J. Bacteriol. 184:5479-5490 (2002).
[3]
RP SEQUENCE FROM N.A.
RC SPECIES=M.bovis; STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12788972;
RA Garnier T., Eidmeyer K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duchoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrall B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882 (2003).
CC -I- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS
CC SUBFAMILY:
CC CAUTION: Ref.l sequence differs from that shown due to a
CC frameshift in position 59. Ref.l sequence has been checked by
CC authors in Ref.1 and they report that no errors have been found.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

DR ENBL; Z73966; CAAS98229.1; ALT_FRAME.
DR ENBL; AE007065; AKK46440.1; -.
DR ENBL; BX248341; CAD96978.1; -.
DR TIGR; MT2159; -.
DR TubercuList; RV2098c; -.
DR InterPro; IPR000084; PE_region.
DR Pfam; PF00934; PE_1.
DR ProDom; PD001223; PE_region_1.
KW Hypothetical protein; Complete proteome.
FT CONFLICT 312 312 G -> GG (IN REF. 1).
SQ SEQUENCE 491 AA; 41979 MW; 12CB630C59CA0C13 CRC64;

Query Match 11.3%; Score 88; DB 1; Length 491;
Best Local Similarity 26.8%; Pred.No. 3 6;
Matches 30; Conservative 13; Mismatches 39; Indels 30; Gaps 5;

QY 20 AGVVPWGGGNN-----HNGGNSSGSPDSTLIYQGSNAALALOSDARKSETTTQS 74
|::|||::|
Db 222 AGLIHGGAGNCGDGCHGSGSKAGSGSGGFQFGGAGGL----- 264
|::|||::|
QY 75 YNGADVDGGAD--NSTLTGTGRNNATIDQWAKNSD---ITVGQYGGN 121
|::|||::|
Db 265 YNGGGAAGSGNGGDAGTGVSDGP---AAGLGSGRGGDAGLIVGGCGGN 313
|::|||::|

RESULT 14
OMB2 NEIMS

RESULT 14
OMB2 NEIMB

ID	OMB2_NEIMB	STANDARD;	PRT;	331 AA.
AC	P30688;			
DT	01-APR-1993 (Rel. 25, Created)			
DT	01-APR-1993 (Rel. 25, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Major outer membrane protein P.IB precursor (Protein IB) (PIB)			
DE	(Porin) (Class 3 protein).			
GN	PORB.			
OS	Neisseria meningitidis (serogroup B).			
OC	Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;			
OC	Neisseriaceae; Neisseria.			
OX	NCBI_TaxID=491;			
RN	[1] _			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CCUG 37604 / M981 / Serogroup B / Serotype 4;			
RX	MEDLINE=CUG 37604; PubMed=1330818;			
RA	Ward M.J., Lambden P.R., Heckels J.E.;			
RT	"Sequence analysis and relationships between meningococcal class 3			
RT	serotype proteins and other porins from pathogenic and non-pathogenic			
RT	Neisseria species.";			
RL	FEMS Microbiol. Lett. 73:283-289(1992).			
RN	[2] _			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CU385 / Serogroup B / Serotype 4 / Subtype 15;			
RX	MEDLINE=93116587; PubMed=1335540;			
RA	Zapata G.A., Vann W.F., Rubinstein Y., Frasch C.E.;			
RT	"Identification of variable region differences in Neisseria			
RT	meningitidis class 3 protein sequences among five group B			
RT	serotypes.";			
RL	Mol. Microbiol. 6:3493-3499(1992).			
CC	-1- FUNCTION: Serves as a slightly cation selective porin.			
CC	-1- SUBUNIT: Homotrimer.			
CC	-1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.			
CC	-1- SIMILARITY: Belongs to the Gram-negative porin family.			
CC	-1- This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	EMBL; X65531; CAA46501.1; -			
DR	PIR; S21409; S21409.			
DR	PIR; S28441; S28441.			
DR	InterPro; IPR001702; Porin Gram-ve.			
DR	Pfam; PF00267; Gram-ve porins; 1.			
DR	PRINTS; PR00182; ECOLNEIPORIN.			
DR	PROSITE; PS00576; GRAM_NEG_PORIN; 1.			
KW	Outer membrane; Porin; Transmembrane; Antigen; Signal.			
FT	SIGNAL 1 19			
FT	CHAIN 20 331			
FT	VARIANT 70 71			
FT	NG -> KR (IN STRAIN CU385).			
FT	SEQUENCE 331 AA; 35741 MW; 35EA35B7EBD28301 CRC64;			
QY	Query Match 11.2%; Score 87; DB 1; Length 331;			
QY	Best Local Similarity 25.8%; Pred. No. 2.8; 78; Indels 22; Gaps 6;			
QY	Matches 42; Conservative 23; Mismatches 23; Gaps 6;			
QY	4 LKVAAPAAIIVSGSALAGVVPQGGGNGHNGGNSGSPDSTLSIYQGS-----A 53			
DB	13 LPVAAMADVTLTYGIRAGV--ETSRVHGNGQVSVETGT-GIVDLGSKIGFKGQEDIG 69			
QY	54 NAALALQSDARKSETTITQSGYCN-----GADVCGQADNSTIETLQNGFRNATIDKNA 108			
DB	70 NGLKAIWQVEKASIAIGTDSGNGRQSFGLKGGFG--KLVRGLNSVLKDTGIDNPMD 127			
QY	109 KNGSDITVQYGGGNAAL--VNYQLVTRVTVTHMAHANATANQY 151			
DB	128 KSDYLGNKTAEPARLISVRYDSPFAGLSGVSQVYALNDNAGKY 172			

RESULT 15	
PER_DROWI	STANDARD; PRT; 1093 AA.
ID	Q03297; O18421; O18422; P91721; P91722;
AC	Q03297; O18421; O18422; P91721; P91722;
DT	01-OCT-1993 (Rel. 27, Created)
DT	15-JUL-1998 (Rel. 36, Last sequence update)
DT	16-OCT-2001 (Rel. 40, Last annotation update)
DE	Period circadian protein (Fragment).
GN	PER.
OS	Drosophila willistoni (Fruit fly).
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC	Ephydroidea; Drosophilidae; Drosophila.
OX	NCBI_TaxID=7260;
RN	[1] _
RP	SEQUENCE FROM N.A.
RC	STRAIN=Various strains;
RX	MEDLINE=97357421; PubMed=9214747;
RA	Gleason J.M., Powell J.R.;
RT	"Interspecific and intraspecific comparisons of the period locus in
RT	the Drosophila willistoni sibling species.";
RL	Mol. Biol. Evol. 14:741-753(1997).
RN	[2] _
RP	SEQUENCE OF 579-646 FROM N.A.
RX	MEDLINE=93196482; PubMed=8450754;
RA	Peixoto A.A., Camposan S., Costa R.H., Kyriacou C.P.;
RT	"Molecular evolution of a repetitive region within the per gene of
RT	Drosophila.";
RL	Mol. Biol. Evol. 10:127-139(1993).
CC	-1- FUNCTION: Essential for biological clock functions. Determines the
CC	period length of circadian and ultradian rhythms; an increase in
CC	PER dosage leads to shortened circadian rhythms and a decrease in
CC	leads to lengthened circadian rhythms. Essential for the circadian
CC	rhythmicity of locomotor activity, eclosion behavior, and for the
CC	thoracic nervous system. The biological cycle depends on the
CC	rhythmic formation and nuclear localization of the TIM-PER
CC	complex. Light induces the degradation of TIM, which promotes
CC	elimination of PER. Nuclear activity of the heterodimer
CC	coordinatively regulates PER and TIM transcription through a
CC	negative feedback loop. Behaves as a negative element in circadian
CC	transcriptional loop. Does not appear to bind DNA, suggesting
CC	indirect transcriptional inhibition (By similarity).
CC	-1- SUBUNIT: Forms heterodimer with timeless (TIM); the complex then
CC	translocates into the nucleus (By similarity).
CC	-1- SUBCELLULAR LOCATION: Nuclear at specific periods of the day.
CC	First accumulates in the perinuclear region about one hour before
CC	translocation into the nucleus. Interaction with Tim is required
CC	for nuclear localization (By similarity).
CC	-1- PTM: PHOSPHORYLATED WITH A CIRCADIAN RHYTHMICITY, PROBABLY BY THE
CC	DOUBLE-TIME PROTEIN (DBT). PHOSPHORYLATION COULD BE IMPLICATED IN
CC	THE STABILITY OF PER MONOMER AND IN THE FORMATION OF HETERODIMER
CC	PER-TIM (BY SIMILARITY).
CC	-1- MISCELLANEOUS: THE SEQUENCE SHOWN IS THAT OF STRAIN ATULXCO.
CC	-1- SIMILARITY: Contains 2 PAS (PER-ARNT-SIM) dimerization domains.
CC	-1- SIMILARITY: Contains 1 PAS-associated C-terminal (PAC) domain.
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	EMBL; U51055; AAB41360.1; -
DR	EMBL; U51056; AAB41361.1; -
DR	EMBL; U51057; AAB41362.1; -
DR	EMBL; U51058; AAB41363.1; -
DR	EMBL; U51059; AAB41364.1; -
DR	EMBL; U51060; AAB41365.1; -
DR	EMBL; U51061; AAB41366.1; -
DR	EMBL; U51062; AAB41367.1; -

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:39:17 ; Search time 29.7 seconds
(without alignments)
1604.150 Million cell updates/sec

Title: US-09-543-407-12

Perfect score: 779

Sequence: 1 MKLLKVAFAAIVVSGSALA.....VTRVVTHEMAHANNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: SP_TREMBL_25.*
- 2: SP_Archea.*
- 3: SP_Bacteria.*
- 4: SP_Fungi.*
- 5: SP_Invertebrate.*
- 6: SP_Mammal.*
- 7: SP_Mbc.*
- 8: SP_Organelle.*
- 9: SP_Phage.*
- 10: SP_Plant.*
- 11: SP_Rodent.*
- 12: SP_Virus.*
- 13: SP_Vertebrate.*
- 14: SP_Unclassified.*
- 15: SP_Virus.*
- 16: SP_Bacteriopl.*
- 17: SP_Archeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	686	88.1	152	2	O33802 salmonella
2	591.5	75.9	150	2	Q7X243
3	537	68.9	149	2	Q7X240
4	495.5	63.6	152	16	Q8CW63
5	427.5	54.9	150	2	Q7X237
6	385	49.4	76	2	Q54069
7	122	15.7	29	2	Q9S3J5
8	113.5	14.6	3501	16	Q8Y106
9	113.5	14.6	3552	16	Q8XSD6
10	108.5	13.9	191	10	Q7XDR3
11	107	13.7	502	16	Q8EIH4
12	106	13.6	1748	5	Q94821
13	105	13.5	2035	2	Q9XCJ4
14	105	13.5	2039	16	Q9ZNS7
15	104.5	13.4	1286	2	Q841Y5
16	104	13.4	2174	16	Q92U08

17	103	13.2	139	16	Q8EIH3
18	102.5	13.2	624	3	Q8NIV1
19	102	13.1	1422	16	Q8EFU3
20	100.5	12.9	152	2	Q7X241
21	100	12.8	179	2	O33801
22	99.5	12.8	151	2	Q7X244
23	99	12.7	1765	16	Q7V8S5
24	98.5	12.6	1222	5	Q9W4F0
25	98.5	12.6	1222	5	Q8T8L9
26	98.5	12.6	1351	5	Q8SX56
27	98.5	12.6	1354	5	Q8MPN4
28	98.5	12.6	7716	16	Q7UWZ8
29	97.5	12.5	154	16	Q89J15
30	97.5	12.5	348	13	Q93397
31	97.5	12.5	739	2	Q9X687
32	97.5	12.5	3659	16	Q98LN6
33	97	12.5	1618	2	Q9KXB1
34	96.5	12.4	151	2	Q7X238
35	96	12.3	145	16	Q8U6N9
36	96	12.3	157	16	Q88HG0
37	95.5	12.3	130	16	Q89J14
38	95.5	12.3	453	5	Q9N6M8
39	95.5	12.3	1615	2	Q9KXA8
40	95	12.2	362	16	Q8EV84
41	95	12.2	1613	2	Q84OU5
42	95	12.2	1618	2	Q9KXB4
43	94.5	12.1	160	16	Q8CW64
44	94.5	12.1	453	5	Q9NGF6
45	94.5	12.1	453	5	Q9NGF7

ALIGNMENTS

RESULT 1

O33802 PRELIMINARY; PRT; 152 AA.

AC O33802; 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE AcfA protein (fragment).
GN AcfA.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98053981; PubMed=9393832;
RA Sukupolvi S.S., Lorentz R.G., Gordon J.I., Bian Z., Pfeifer J.D.,
RA Normark S.J., Rhen M.;
RT "Expression of thin, aggregative fimbriae promotes interaction of
RT Salmonella typhimurium 8R-11 with mouse small intestinal epithelial
RT cells."
RL Infect. Immun. 65:5320-5325(1997).
DR ENBL, AJ000514; CAA04151.1; .
FT NON_TER 152 152
SQ SEQUENCE 152 AA; 15401 MW; 9DA7DADC2364B006 CRC64;

Query Match 88.1%; Score 686; DB 2; Length 152;

Best Local Similarity 88.7%; Pred. No. 4.1e-45;

Matches 134; Conservative 5; Mismatches 12; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGNGSGSGPDSSTLSIVQYGSANAALAIQ 60

Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGNGSGSGPDSSTLSIVQYGSANAALAIQ 60

QY 61 SPARKSETTITSGGNGADVGCGADNSTIETQNGFRNNATIDOWNAKNSDITVQYGG 120

Db 61 SPARKSETTITSGGNGADVGCGADNSTIETQNGFRNNATIDOWNAKNSDITVQYGG 120

QY 121 NNAALVNDQLVTRVVTHEMAHANNATANQY 151

```

Db      121 NNAALVNTQASDSSVVMVQVGFNNAPANQY 151
      ||||| : : : ||| |||
      121 NNAALVNTQASDSSVVMVQVGFNNAPANQY 151

RESULT 2
Q7X243
ID      Q7X243      PRELIMINARY;      PRT;      150 AA.
AC      Q7X243;
DT      01-OCT-2003 (TReMBLrel. 25, Created)
DT      01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DT      01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE      Curlin-csgA protein.
GN      CSGA.
OS      Citrobacter sp. Fec2.
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC      Enterobacteriaceae; Citrobacter.
OX      NCBI_TaxID=213763;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=Fec2;
RA      Zogaj X., Bokranz W., Nimitz M., Romling U.;
RT      "Production of Cellulose and Curli Fimbriae by Members of the Family
RT      Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL      Infect. Immun. 72:4151-4158(2003).
DR      EMBL; AJ515700; CAD56672.1; -.
SQ      SEQUENCE 150 AA; 15016 MW; 1D7141B8D6973DC6 CRC64;

Query Match      75.9%; Score 591.5; DB 2; Length 150;
Best Local Similarity 77.5%; Pred. No. 6.7e-38;
Matches 117; Conservative 14; Mismatches 19; Indels 1; Gaps 1;

QY      1 MKLLKVAAPAAIIVVSGSALAGVVPQW-GGNGHGGSNVGPSSLSIYQYGSNANALQ 60
DB      1 MKLLQVAAPAAIIVVSGSALAGVVPQWGGGG-GGGSSSGSPSTLSIYQYGSNANALQ 59
QY      61 SDARKSETTITQSGYNGADVGGADNSTIELTQNGFRNNAIDQWNAKNSDITVGYGG 120
DB      60 SDARKSETTITQNGFGNGADVGGSDNSTIDITQNGFKNAIDQWNGKNSDITVSYGG 119

QY      121 NNAALVNDLVTRVVTHEMAHANNATANQY 151
DB      120 HNAALVNTQASDSSVLVQVGFNNATANQY 150

RESULT 3
Q7X240
ID      Q7X240      PRELIMINARY;      PRT;      149 AA.
AC      Q7X240;
DT      01-OCT-2003 (TReMBLrel. 25, Created)
DT      01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DT      01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE      Curlin-csgA protein.
GN      CSGA.
OS      Citrobacter freundii.
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC      Enterobacteriaceae; Citrobacter.
OX      NCBI_TaxID=546;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=fec4;
RA      Zogaj X., Bokranz W., Nimitz M., Romling U.;
RT      "Production of Cellulose and Curli Fimbriae by Members of the Family
RT      Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL      Infect. Immun. 72:4151-4158(2003).
DR      EMBL; AJ515701; CAD56675.1; -.
SQ      SEQUENCE 149 AA; 15260 MW; 946DD52017F648FD CRC64;

Query Match      68.9%; Score 537; DB 2; Length 149;
Best Local Similarity 71.5%; Pred. No. 9.7e-34;
Matches 108; Conservative 17; Mismatches 24; Indels 2; Gaps 1;

QY      1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGNGHGGSNVGPSSLSIYQYGSNANALQ 60

```

RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RT Enterobacteriaceae Isolated from the Human Gastrointestinal Tract."
RL Infect. Immun. 72:4151-4158(2003).
DR EMBL: AJ515702; CAD56678.1; -
SQ SEQUENCE 150 AA; 15112 MW; 506BB2D872DF15F3 CRC64;

Query Match 54.9%; Score 427.5; DB 2; Length 150;
Best Local Similarity 58.3%; Pred. No. 2.3e-25;
Matches 88; Conservative 26; Mismatches 36; Indels 1; Gaps 1;

QY 1 MKLLKVAFAAIVVSGSALAGVPCWGGGNGHNGSSGPDSTLSIYQYGSANAALQ 60
DB 1 MKFIKVAALAAIVVSGSAGMAGNIQ-GGWHGHGGHGGGPNSTLNIYQNGGNSALQ 59

QY 61 SDARKSETTITQSGYNGADYQGGADNSTIELTQNGFRNNAITDQWNAKNSDIIVGYGG 120
DB 60 TDARNSVLNISQTEGGGADYQGGSDSSINLTQNGFNSATLQWNSKDSVMNVSYGG 119

QY 121 NNAALVNDQLVTRVTHEMAHANNATNOY 151
DB 120 LNALVDQTSNSTVNTQIGFGNHATAHQY 150

RESULT 6
Q54069 PRELIMINARY; PRT; 76 AA.
AC Q54069;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE SEF17 fimbria (Fragment).
GN AGFA.
OS Salmonella enteritidis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=592;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SE30;
RA Cox J.M., Eglezos S., Woolcock J.B.;
RT "Virulence of Salmonella enteritidis in chickens correlates with
RT colony morphology and expression of SEF17 fimbriae."
RL Submitted (Apr-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: U53207; AAA98671.1; -
FT NON_TER 1
FT NON_TER 76
SQ SEQUENCE 76 AA; 7704 MW; 2FD5411241A7BCB1 CRC64;

Query Match 49.4%; Score 385; DB 2; Length 76;
Best Local Similarity 98.7%; Pred. No. 1.9e-22;
Matches 75; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 30 GNHNGGSSGPDSTLSIYQYGSANAALQSDARKSETTITQSGYNGADYQGGADNST 89
DB 1 GNHNGGSSGPDSTLSIYQYGSANAALQSDARKSETTITQSGYNGADYQGGADNST 60

QY 90 IELTQNGFRNNAITDQ 105
DB 61 IELTQNGFRNNAITDQ 76

RESULT 7
Q9S3J5 PRELIMINARY; PRT; 29 AA.
AC Q9S3J5;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-MAY-2001 (TRENBLrel. 16, Last annotation update)
DE Curlin subunit monomer (Fragment).
GN CSGA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.

OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC TRANSPOSON-insertion sequence IS1;
RX MEDLINE=99314153; PubMed=10386375;
RA La Ragione R.M., Collighan R.J., Woodward M.J.;
RT "Non-cultivation of Escherichia coli O78:H80 isolates associated with
RT ISI insertions in csgH and reduced persistence in poultry infection."
RL FEMS Microbiol. Lett. 175:247-253(1999).
DR EMBL: AJ131756; CAB45380.1; -
FT NON_TER 29
SQ SEQUENCE 29 AA; 2789 MW; E290DFC07ABBB243 CRC64;

Query Match 15.7%; Score 122; DB 2; Length 29;
Best Local Similarity 89.7%; Pred. No. 0.0078; 2; Indels 0; Gaps 0;
Matches 26; Conservative 1; Mismatches 1;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGG 29
DB 1 MKLLKVAFAAIVVSGSALAGVVPQYGGG 29

RESULT 8
Q8Y106 PRELIMINARY; PRT; 3501 AA.
AC Q8Y106;
DT 01-MAR-2002 (TRENBLrel. 20, Created)
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Probable hemagglutinin-related protein.
GN RSC0887 OR RSC06116.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GMI1000;
RX MEDLINE=21691879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Arlat M., Billault A., Brotier P., Camus J.C., Catolico L.,
RA Chandler M., Choisme N., Claudel-Renard C., Cunnac S., Denange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schliex F.,
RA Siguer P., Thebault P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum."
RL Nature 415:497-502(2002).
DR EMBL: AF646061; CAD14589.1; -
DR GO: GO:0004519; F:Endonuclease activity; IEA.
DR GO: GO:0003676; F:Nucleic acid binding; IEA.
DR InterPro: IPR001604; Endonuclease.
DR InterPro: IPR008619; Fil haemagg.
DR InterPro: IPR008638; Haemagg act.
DR Pfam: PF05594; Fil haemagg act.
DR Pfam: PF05860; Haemagg act; 1.
DR PROSITE: PS01070; NUCLEASE_NON_SPEC; 1.
SQ Complete proteome.
KW SEQUENCE 3501 AA; 348421 MW; 290B41C99018A107 CRC64;

Query Match 14.6%; Score 113.5; DB 16; Length 3501;
Best Local Similarity 28.8%; Pred. No. 8.3;
Matches 36; Conservative 19; Mismatches 43; Indels 27; Gaps 4;

QY 15 SGSALAGVVPQWGGGNGHNGG-NSSGPDSTLSIYQYGSANAA----- 56
DB 2431 SGSHFSTAGPSWGLGRNVGGGPNSSG---VGLAFYGSASADNAAGNSRQNASVWIG 2486

QY 57 LALQSDARKSETTITQSGYNGADYQGGADNSTIELTQNGFRNNAITDQWNAKNSDIIVG 116
DB 2487 KSVQVQARTGDIIVVSGSALSDVLLAKQKQVDIVAGNDTSSRHD-----HSDRTIG 2541

QY 117 QYGN 121
|||

```

Db      2542 DLGGN 2546

RESULT 9
QBXSDB
ID      QBXSDB PRELIMINARY; PRT; 3552 AA.
AC      QBXSDB;
DT      01-MAR-2002 (TrEMBLrel. 20, Created)
DT      01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT      01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE      Probable hemagglutinin-related protein.
GN      RSP0540 OR RS06117.
OS      Ralstonia solanacearum (Pseudomonas solanacearum).
OC      Plasmid megaplasmid.
OC      Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC      Burkholderiaceae; Ralstonia.
OX      NCBI_TaxID=305;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=GM11000;
RX      MEDLINE=21681879; PubMed=11823852;
RA      Salanoubat M., Genin S., Arriguenave F., Gouzy J., Mangenot S.,
RA      Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA      Chandler M., Choise N., Claudel-Renard C., Cunnac S., Demange N.,
RA      Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
RA      Siguer P., Thebault P., Whalen M., Wincker P., Levy M.,
RA      Weissenbach J., Boucher C.A.;
RT      "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL      Nature 415:497-502(2002).
DR      EMBL: AL646079; CAD17691.1; -
DR      GO: 0046821; C:extrachromosomal DNA; IEA.
DR      GO: 0004519; F:endonuclease activity; IEA.
DR      GO: 0003676; F:nucleic acid binding; IEA.
DR      InterPro: IPR001604; Endonuclease.
DR      InterPro: IPR008619; Fil haemagg.
DR      InterPro: IPR008638; Haemagg act.
DR      Pfam: PF05594; Fil haemagg; 20.
DR      Pfam: PF05860; Haemagg act; 1.
DR      PROSITE: PS01070; NUCLEASE_NON_SPEC; 1.
KW      Plasmid; Complete proteome.
SQ      SEQUENCE 3552 AA; 352934 MW; C5432AABE2CCF59C CRC64;

Query Match 14.6%; Score 113.5; DB 16; Length 3552;
Best Local Similarity 28.8%; Pred. No. 8.4;
Matches 36; Conservative 19; Mismatches 43; Indels 27; Gaps 4;

QY      15 GSalAGVVPQGGGNNHGGG-NSGPDSTLSIYQGSANA-----56
Db      2430 SGSHFTAGTSGWDLGRNVGGGSSG---VGLAPYGSASDAAGNSRQNASVVG 2485

QY      57 LALQSDARKSETITQSGYNGADYCGQADNSTIETQGFNNATIDQNAKNSDITVG 116
Db      2486 KSVQVQARTGDTIVSGSGISALSDVLLAKQKGVIVAGNDSRSHED-----HSDRTIG 2540

QY      117 QYGGN 121
Db      2541 DLGGN 2545

RESULT 10
Q7XDR3
ID      Q7XDR3 PRELIMINARY; PRT; 191 AA.
AC      Q7XDR3;
DT      01-OCT-2003 (TrEMBLrel. 25, Created)
DT      01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT      01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE      Putative glycine-rich protein.
GN      OSJNB0039P10.07G10.
OS      Oryza sativa (japonica cultivar-group).
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC      Ehrhartoideae; Oryzaceae; Oryza.
OX      NCBI_TaxID=39947;

[1]
RN      SEQUENCE FROM N.A.
RP      STRAIN=cv. Nipponbare;
RC      STRAIN=cv. Nipponbare;
RA      Buell C.R., Wing R.A., McCombie W.R., Messing J., Yuan Q.;
RA      Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
RL      EMBL: AE017101; AAP54078.1; -.
DR      SEQUENCE 191 AA; 17021 MW; 95A246FE11DE3AC5 CRC64;

Query Match 13.9%; Score 108.5; DB 10; Length 191;
Best Local Similarity 29.5%; Pred. No. 0.72; 59; Indels 5; Gaps 2;
Matches 33; Conservative 15; Mismatches 15; Indels 28; Gaps 4;

QY      16 GSALAGVVPQGGGNNHGGGNSGPDSTL--SIYQGSANAALALQSPARKSETITQS 73
Db      72 GGGGGGGYGYGGSGSGSGSSQTSQNGYGYGGSSAGSGAGAGAGAGAGGYP 131

QY      74 GYNGADVGQADNSIETQGFNNATIDQNAKNSDITVQYGGNNAAL 125
Db      132 SYGHGSGSGTSGSSEATNTWYQYANAGNGGNGGNGQ--GQYGGSGAGV 190

RESULT 11
Q8EIH4
ID      Q8EIH4 PRELIMINARY; PRT; 502 AA.
AC      Q8EIH4;
DT      01-MAR-2003 (TrEMBLrel. 23, Created)
DT      01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT      01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE      Conserved hypothetical protein.
GN      S00865.
OS      Shewanella oneidensis.
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC      Alteromonadaceae; Shewanella.
OX      NCBI_TaxID=70863;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=MB-1.
RX      MEDLINE=22297686; PubMed=12168813;
RA      Heidelberg J.F., Paulsen I., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA      Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
RA      Meyer T., Tsapin A., Scott J., Beaman M., Brinkac L., Daugherty S.,
RA      DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA      Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
RA      Vamathevan J., Weidman J., Imbraim M., Lee K., Berry K., Lee C.,
RA      Mueller J., Khouri H., Gill J., Utterback T.R., McDonald L.A.,
RA      Feldblyum T.V., Smith H.O., Venter J.C., Nelson K.H., Fraser C.M.;
RT      "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT      Shewanella oneidensis.";
RL      Nat. Biotechnol. 20:1118-1123(2002).
DR      EMBL: AE015532; AAN53941.1; -.
DR      TIGR: S00865; -.
KW      Hypothetical protein; Complete proteome.
SQ      SEQUENCE 502 AA; 52441 MW; D09CA23D6C46B62D CRC64;

Query Match 13.7%; Score 107; DB 16; Length 502;
Best Local Similarity 27.3%; Pred. No. 2.8;
Matches 33; Conservative 15; Mismatches 45; Indels 28; Gaps 4;

QY      29 GGNHNG-----GGN-----SSGPDSTLSIYQGSANA---ALALQS 61
Db      231 GDNHTGTVYALAGSENDISMEQGSNTAYLSMTTGDNTVDTIQDGSNTVGDLSIADI 290

QY      62 DARKSETTITQSGYNGADYCGQADNSTIETQGFNNATIDQNAKNSDITVQYGGN 121
Db      291 QGDDNDITIKQKGSNGAEFQVWGDSDVDLQKGDANFATFGAYGTDN--DFDLSSKGDN 349

```



```
QY 122 N 122
Db 350 N 350

RESULT 12
Q94821 PRELIMINARY; PRT; 1748 AA.
ID Q94821 P92146; P92145; P92144; P92143; P92142; P92141; Q94820;
AC Q94821 P92146; P92145; P92144; P92143; P92142; P92141; Q94820;
DT 01-FEB-1997 (TREMELrel. 02, Created)
DT 01-FEB-1997 (TREMELrel. 02, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE CNUB protein.
GN CNUB.
OS Tetrahymena thermophila.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
OC Tetrahymenina; Tetrahymena.
OX NCBI_TaxID=5911;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88189811; PubMed=3357771;
RA Martindale D.W., Taylor F.M.;
RT "Multiple introns in a conjugation-specific gene from Tetrahymena thermophila."
RT Nucleic Acids Res. 16:2189-2201(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94051569; PubMed=8233798;
RA Taylor F.M., Martindale D.W.;
RT "Retroviral-type zinc fingers and glycine-rich repeats in a protein encoded by cnjB, a Tetrahymena gene active during meiosis."
RT Nucleic Acids Res. 21:4610-4614(1993).
DR EMBL; X06462; CAB37323.1; -.
DR EMBL; L03710; AAC37171.1; -.
DR PIR; S42136; S42136.
DR HSSP; P05888; 1A4F.
GO; GO:0003676; P:nucleic acid binding; IEA.
DR InterPro; IPR001878; Znf CCHC.
DR Pfam; PF00098; Zf-CCHC; 7.
DR PRINTS; PR00939; C2HCZNFINGER.
DR SMART; SMO0343; Znf_C2HC; 7.
DR PROSITE; PS0158; ZF_CCHC; 7.
FT CONFLICT 251 251 M -> I (IN REF. 1).
FT CONFLICT 256 256 I -> N (IN REF. 1).
SQ SEQUENCE 1748 AA; 199624 MW; 0B03F210104008A3 CRC64;

Query Match 13.6%; Score 106; DB 5; Length 1748;
Best Local Similarity 32.1%; Pred. No. 14;
Matches 35; Conservative 13; Mismatches 33; Indels 28; Gaps 5;

QY 25 QWGGGNGHGG---GNSSGPDSTLSIYQYGSANAALALQSPARKSETTIT---QSGVGN 77
Db 1640 QFGGGNGNGGSGWGTSSGSDN-----CQSNVQSESTTSSGGWSSGSGN 1685

QY 78 GADVQGGADNSTIETLTQNGFRNNATIDOWNAKNSDITVGY--YGGNNAA 124
Db 1686 QTGGGWSGNDN-----QQQENNTGGGWSGNSQNTNNESSGWSNNQA 1729

RESULT 13
Q9XCJ4 PRELIMINARY; PRT; 2035 AA.
ID Q9XCJ4
AC Q9XCJ4
DT 01-NOV-1999 (TREMELrel. 12, Created)
DT 01-NOV-1999 (TREMELrel. 12, Last sequence update)
DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
DE ShdA.
GN SHDA.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC14028;
RA Kingsley R.A., van Amsterdam K., Baumber A.J.;
RT "The presence of a pathogenicity island specific to Salmonella enterica subspecies I correlates with adaptation to warm blooded animals."
RT Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC14028;
RA Kingsley R.A., van Amsterdam K., Edwards E.W., Hargis B.M., Baumber A.J.;
RT "Complete sequence of the xseA-his intergenic region of the S. enterica serotype Typhimurium genome and its distribution within the genus Salmonella."
RT Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
DR EMBL; A1140550; AAD25110.2; -.
DR InterPro; IPR006315; Autotransport.
DR InterPro; IPR005546; Autotransporter.
DR InterPro; IPR004899; Pertactin.
DR InterPro; IPR002173; PfkB.
DR Pfam; PF03797; Autotransporter; 1.
DR Pfam; PF03212; Pertactin; 1.
DR TIGRFAMs; TIGR01414; autotrans_bar1; 3.
DR PROSITE; PSC0584; PFKB_KINASES_2; 2.
SQ SEQUENCE 2035 AA; 207032 MW; 295DB82FFCA84FAB CRC64;

Query Match 13.5%; Score 105; DB 2; Length 2035;
Best Local Similarity 26.3%; Pred. No. 20;
Matches 54; Conservative 20; Mismatches 61; Indels 70; Gaps 11;

QY 10 AAIYVSGSALAG-----VVPQGGGNGHNGGG---NSSGPD----- 42
Db 90 AAIYVSGVATVWQPTTGTGLVETSGGADDPGGKYVSNALSDHVAILELTDAKI 149
QY 43 STLSIYQYGSANA-----ALALQSPARKS-ETTITQSGVGNADVG 82
Db 150 TTTGIYTGISAAAGSTLRLDSTLTIDNGFVMTLYTGTSEATLDGTIVEAANSSAQVQ 209

QY 83 QGA-----DNSTIELTQ---NGFRNNATIDOWNAKN-SDITVGYQY-----GNNAAALVN 127
Db 210 QGSTNLVLDGSTLTITLAQQINNVAGTATDEGSTLNLSDSSVSSAGTMTIOGTNKALN 269

QY 128 YDQLVTRVVTHEMA-----HANNAT 147
Db 270 ---LTNATITHNASCAGAAVQANNAT 291

RESULT 14
Q8ZN57 PRELIMINARY; PRT; 2039 AA.
ID Q8ZN57
AC Q8ZN57;
DT 01-MAR-2002 (TREMELrel. 20, Created)
DT 01-MAR-2002 (TREMELrel. 20, Last sequence update)
DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
DE Similar to the C-terminal region of AIDA, IcsA, subspecies I specific, Peyer's patch colonization and shedding factor.
DE SHDA OR STM2513.
GN Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Lt2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P., Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D., Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E., Van E., Sun H., Florea L., Miller W., Stoneking T., Nhan M., Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium Lt2.";
```

```

RL Nature 413:852-856 (2001).
DR EMBL: AEO08813; AAL21407.1; -.
DR InterPro: IPR006315; Autotransport.
DR InterPro: IPR005546; Autotransporter.
DR InterPro: IPR004899; Pertactin.
DR InterPro: IPR002173; PfkB.
DR Pfam: PF03797; Autotransporter; 1.
DR TIGRFAMs: TIGR01414; autotrans_barl; 3.
DR PROSITE: PS00584; PFKB_KINASES_2; 2.
KW Complete proteome.
SQ SEQUENCE 2039 AA; 207127 MW; 894E41F8F29339EA CRC64;

Query Match 13.5%; Score 105; DB 16; Length 2039;
Best Local Similarity 26.3%; Pred. No. 20;
Matches 54; Conservative 20; Mismatches 61; Indels 70; Gaps 11;

QY 10 AAIIVSGSALAG-----VVPQGGGGNHNHGGG-----NSSGPD----- 42
DB 94 AALYVSGVATVGQPTTIVTGTGLVETSGGADDPGGKYVSNALSLDHYAILELTDAKI 153
QY 43 STLISIVQGSANA-----ALALQSDARKS-ETTITQSGYNGADV 82
DB 154 TTGCIYQGISADGSLTBLTSTLTIDNFGVMTLYTGEATLDGTIVEANSSAQVQ 213

QY 83 QGA-----DNSTIELTQ---NGFRNNATIDQNAKN-SDITVGYQG-----GNNAALVN 127
DB 214 QGSTNLVLDGSTITLAQQGQINNVAGNTATDEGSTLNLSDSVSSAGTMTSTIQGTNKAALN 273

QY 128 YDQIVTRVVTHEMA-----HANNAT 147
DB 274 ---LTWATITHNAGAAVQANAT 295

RESULT 15
Q841Y5 PRELIMINARY; PRT; 1286 AA.
ID Q841Y5;
AC Q841Y5;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DE Putative high-molecular-weight surface-exposed protein Cf0009.
OS Campylobacter fetus.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Campylobacteraceae; Campylobacter.
OX NCBI_TaxId=196;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=23D;
RX MEDLINE=90354448; PubMed=2387868;
RA Blaser M.J., Gotschlich E.C.;
RT "Surface array protein of Campylobacter fetus. Cloning and gene
structure.";
RL J. Biol. Chem. 265:14529-14535 (1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=23D;
RX MEDLINE=9103477; PubMed=2229082;
RA Blaser M.J., Gotschlich E.C.;
RT "Surface array protein of Campylobacter fetus. Cloning and gene
structure.";
RL J. Biol. Chem. 265:19372-19372 (1990).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=23D;
RX MEDLINE=92394895; PubMed=1522068;
RA Tummuru M.K., Blaser M.J.;
RT "Characterization of the Campylobacter fetus sapA promoter: evidence
that the sapA promoter is deleted in spontaneous mutant strains.";
RL J. Bacteriol. 174:5916-5922 (1992).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=23D;
MEDLINE=93348254; PubMed=8346244;
RA Tummuru M.K., Blaser M.J.;
RT "Rearrangement of sapA homologs with conserved and variable regions in
Campylobacter fetus.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:7265-7269 (1993).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=23D;
RX MEDLINE=95204338; PubMed=7896695;
RA Dworkin J., Tummuru M.K., Blaser M.J.;
RT "A lipopolysaccharide-binding domain of the Campylobacter fetus S-
layer protein resides within the conserved N terminus of a family of
silent and divergent homologs.";
RL J. Bacteriol. 177:1734-1741 (1995).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=23D;
RX MEDLINE=99059317; PubMed=9851986;
RA Thompson S.A., Shedd O.L., Ray K.C., Beins M.H., Jorgensen J.P.,
Blaser M.J.;
RT "Campylobacter fetus surface layer proteins are transported by a type
I secretion system.";
RL J. Bacteriol. 180:6450-6458 (1998).
RN [7]
RP SEQUENCE FROM N.A.
RC STRAIN=23D;
RX PubMed=12694614;
RA Tu Z.C., Wassenaar T.M., Thompson S.A., Blaser M.J.;
RT "Structure and genotypic plasticity of the Campylobacter fetus sap
locus.";
RL Mol. Microbiol. 48:685-698 (2003).
DR EMBL: AY211269; AAC64216.1; -.
DR InterPro: IPR005546; Autotransporter.
DR Pfam: PF03797; Autotransporter; 1.
SQ SEQUENCE 1286 AA; 134079 MW; A1FF9CBG34:58789 CRC64;

Query Match 13.4%; Score 104.5; DB 2; Length 1286;
Best Local Similarity 27.8%; Pred. No. 13;
Matches 44; Conservative 26; Mismatches 57; Indels 31; Gaps 9;

QY 5 KVAFAAIVVSGSALAGVVPQGGGGNHNHGGNS---SGPDTLSIYQYGSANAALQOS 61
DB 376 QVASENLVVISGGIN--VPTIGGGSATWATNNQVTISGCKVTSTIYGCNAN-----K 427
QY 62 DARKSETTITQSGYNGADV--GQAGDNSTIELTQNGFNATIDQNAKNSDITVGYQG 119
DB 428 SANENKVITE-GTANVADIYGGKSISSNSI-----ANKNSITISGGTLQVTNI----YG 477
QY 120 GNNAALVNYDQL-----VTRVV-THEMAHANNATAN 149
DB 478 GHSKADANENSIQISNGGNNINNVGGHAQDHTNLNTIN 515

Search completed: August 2, 2004, 14:54:36
Job time : 31.7 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:35:42 ; Search time 44.9 seconds

(without alignments)
950.215 Million cell updates/sec

Title: US-09-543-407-14

Perfect score: 775

Sequence: 1 MKLKVAFAAIVVSGSALA.....HASVWVQVGFNNATANQY 151

Scoring table: BL0SUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseqp_29Jan04.*

1: Geneseqp1980s.*

2: Geneseqp1990s.*

3: Geneseqp2000s.*

4: Geneseqp2001s.*

5: Geneseqp2002s.*

6: Geneseqp2003as.*

7: Geneseqp2003bs.*

8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	775	100.0	151	3	AAB36347	Aab36347 Agfa::PT3
2	714	92.1	151	3	AAB36352	Aab36352 Agfa::PT3
3	696	89.8	151	3	AAB36346	Aab36346 Agfa::PT3
4	693	89.4	151	2	AAR74625	Aar74625 Agfa sequ
5	693	89.4	151	3	AAB36341	Aab36341 Salmonell
6	692	89.3	151	2	AAW23570	Aaw23570 Salmonell
7	659	85.0	151	3	AAB36355	Aab36355 Agfa::PT3
8	612	79.0	151	3	AAB36353	Aab36353 Agfa::PT3
9	611	78.8	151	3	AAB36349	Aab36349 Agfa::PT3
10	609	78.6	151	3	AAB36350	Aab36350 Agfa::PT3
11	603	77.8	151	3	AAB36354	Aab36354 Agfa::PT3
12	602	77.7	151	3	AAB36351	Aab36351 Agfa::PT3
13	578	74.6	151	3	AAB36348	Aab36348 Agfa::PT3
14	528	68.1	151	3	AAB36343	Aab36343 Escherich
15	523	67.5	151	7	ABR82651	ABr82651 E. coli C
16	514	66.3	120	2	AAR62761	Aar62761 Agfa sequ
17	514	66.3	120	2	AAW23569	Aaw23569 Salmonell
18	463	59.7	142	2	AAR52664	Aar52664 Fibronect
19	391	50.5	122	2	AAR52663	Aar52663 FNB curli
20	237	30.6	45	3	AAB36316	Aab36316 Salmonell
21	132	17.0	22	3	AAB36318	Aab36318 Salmonell
22	123	15.9	23	3	AAB36321	Aab36321 Salmonell
23	123	15.9	23	3	AAB36326	Aab36326 Salmonell
24	123	15.9	23	3	AAB36338	Aab36338 Salmonell
25	115	14.8	22	3	AAB36325	Aab36325 Salmonell

26	115	14.8	22	3	AAB36339	Aab36339 Salmonell
27	115	14.8	22	3	AAB36320	Aab36320 Salmonell
28	113	14.6	24	7	ABR82644	ABr82644 E. coli C
29	109	14.1	23	3	AAB36340	Aab36340 Salmonell
30	109	14.1	23	3	AAB36324	Aab36324 Salmonell
31	109	14.1	23	3	AAB36319	Aab36319 Salmonell
32	104.5	13.5	151	3	AAB36344	Aab36344 Escherich
33	98.5	12.7	151	3	AAB36342	Aab36342 Salmonell
34	98	12.6	26	7	ABR82645	ABr82645 E. coli C
35	95	12.3	186	6	ABU21488	ABu21488 Protein e
36	93	12.0	19	3	AAB36323	Aab36323 Salmonell
37	93	12.0	19	3	AAB36336	Aab36336 Salmonell
38	93	12.0	19	3	AAB36328	Aab36328 Salmonell
39	92	11.9	23	3	AAB36331	Aab36331 Escherich
40	92	11.9	673	3	AAV44403	Aay44403 Human tru
41	92	11.9	673	5	AAU79538	Aau79538 Truncated
42	92	11.9	949	3	AAV44404	Aay44404 Human tru
43	92	11.9	949	5	AAU79539	Aau79539 Truncated
44	92	11.9	1327	3	AAV44402	Aay44402 Human tan
45	92	11.9	1327	3	AAB27212	Aab27212 Human tan

ALIGNMENTS

RESULT 1
AAB36347
ID AAB36347 standard; protein; 151 AA.
XX
AC AAB36347;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa::PT3#2 amino acid sequence SEQ ID NO:14.
XX
KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
XX vaccine; immune response; immunogen.
XX OS Salmonella enteritidis.
XX OS Escherichia coli.
XX OS Synthetic.
XX PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UVVI-) UNIV VICTORIA.
XX
FI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
XX
PS N-PSDB; AAC64623.
XX
PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
XX which encodes foreign epitope or antigen, expresses recombinant agfa
XX protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 136; 139pp; English.

The present invention describes a recombinant agfa gene (I) where a segment of the gene has been replaced by a segment of a foreign DNA sequence which encodes a foreign epitope or antigen. Also described are: (1) use of thin aggregative fimbriae (SEF7/7AF) nucleation depended assembly system of strains of Salmonella, Escherichia coli and Enterobacteriaceae for the production of fimbriae comprising recombinant Agfa, CsgA and Agfa-homologue fimbrin subunits, respectively; (2) directing recombination of a recombinant gene into the chromosome of the homologous species; (3) directing recombination of a recombinant gene back into the chromosome of the homologous species, replacing the native

CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbriae protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;

Query Match 100.0%; Score 775; DB 3; Length 151;
 Best Local Similarity 100.0%; Pred. No. 5.4e-67;
 Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MKLLKVAAPAAIIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
 DB 1 MKLLKVAAPAAIIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
 QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGQYDQ 120
 DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGQYDQ 120
 QY 121 LVTRVVTHEMAHASVMVRQVGFNNATANQY 151
 DB 121 LVTRVVTHEMAHASVMVRQVGFNNATANQY 151

RESULT 2
 AAB36352
 ID AAB36352 standard; protein; 151 AA.
 XX
 AC AAB36352;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Agfa::PT3#7 amino acid sequence SEQ ID NO:24.
 XX
 KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
 XX vaccine; immune response; immunogen.
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 PN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay WW;
 XX WPI; 2000-672631/65.
 DR N-PSDB; AAC64628.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 XX protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 138; 139pp; English.

XX The present invention describes a recombinant agfa gene (J) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/7AF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC -Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbriae protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;

Query Match 92.1%; Score 714; DB 3; Length 151;
 Best Local Similarity 91.1%; Pred. No. 4.3e-61;
 Matches 144; Conservative 0; Mismatches 0; Indels 14; Gaps 2;
 QY 1 MKLLKVAAPAAIIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
 DB 1 MKLLKVAAPAAIIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
 QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGQYDQ 120
 DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGQYDQ 113
 QY 121 LVTRVVTHEMAHASVMVRQVGFNNATANQY 151
 DB 121 LVTRVVTHEMAHASVMVRQVGFNNATANQY 151

RESULT 3
 AAB36346
 ID AAB36346 standard; protein; 151 AA.
 XX
 AC AAB36346;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Agfa::PT3#1 amino acid sequence SEQ ID NO:12.
 XX
 KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
 XX vaccine; immune response; immunogen.
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 PN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.

XX White AP, Doran JL, Collison SK, Kay WW;
 XX WPI; 2000-672631/65.
 DR N-PSDB; AAC64622.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 135; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (Sf17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 XX Sequence 151 AA;
 Query Match 89.8%; Score 696; DB 3; Length 151;
 Best Local Similarity 87.6%; Pred. No. 2.4e-59;
 Matches 141; Conservative 0; Mismatches 0; Indels 20; Gaps 2;
 QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQVGSANAALALQ 60
 DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQVGSANAALALQ 60
 QY 61 SDARKSETTITQSGYNGADVGQGDNSTIETQNGFRNNATIDQWNAKNSDITVQYQG 117
 DB 61 SDARKSETTITQSGYNGADVGQGDNSTIETQNGFRNNATIDQWNAKNSDITVQYQG 120
 QY 118 -----YDLVTRVVTTHMAHASVMVROVGFNNATANQY 151
 DB 121 NNAALVNYDQVTRVVTTHMAHA-----NNATANQY 151
 RESULT 4
 AAR74625
 ID AAR74625 standard; protein; 151 AA.
 XX
 AC AAR74625;
 XX
 XX 25-MAR-2003 (revised)
 DT 26-JUN-1995 (first entry)
 XX
 XX Agfa sequence.
 XX
 XX Salmonella; Agfa; vaccine.
 KW White AP, Doran JL, Collison SK, Kay WW;
 OS WPI; 2000-672631/65.
 XX

PN WO9425598-A2.
 XX
 XX 10-NOV-1994.
 XX
 XX 26-APR-1994; 94WO-IB000207.
 XX
 XX 26-APR-1993; 93US-00054452.
 XX
 XX (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
 PA (KING/) KING J.
 XX
 XX Kay WW, Collinson SK, Clouthier SC, Doran JL;
 PI
 XX WPI; 1994-358275/44.
 DR N-PSDB; AAQ37467.
 XX
 XX Eliciting an immune response to Salmonella - using attenuated Salmonella
 PT strains, vector constructs, or compsns. contg. fimbrial type proteins.
 XX
 XX Disclosure; Fig 7B; 95pp; English.
 PS
 XX The Salmonella Agfa protein and DNA are used in vaccine and genetic
 CC immunization compositions, respectively, to elicit an immune response to
 CC Salmonella in animals (e.g. food producing animals) and humans. (Updated
 CC on 25-MAR-2003 to correct PN field.)
 XX
 XX Sequence 151 AA;
 Query Match 89.4%; Score 693; DB 2; Length 151;
 Best Local Similarity 90.7%; Pred. No. 4.7e-59;
 Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;
 QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQVGSANAALALQ 60
 DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQVGSANAALALQ 60
 QY 61 SDARKSETTITQSGYNGADVGQGDNSTIETQNGFRNNATIDQWNAKNSDITVQYQG 120
 DB 61 SDARKSETTITQSGYNGADVGQGDNSTIETQNGFRNNATIDQWNAKNSDITVQYQG 120
 QY 121 LVTRVVTTHMAHASVMVROVGFNNATANQY 151
 DB 121 NNAALVNYDQVTRVVTTHMAHA-----NNATANQY 151
 RESULT 5
 AAB36341
 ID AAB36341 standard; protein; 151 AA.
 XX
 AC AAB36341;
 XX
 XX 26-FEB-2001 (first entry)
 DT
 XX
 XX Salmonella enteritidis Agfa amino acid sequence SEQ ID NO:5.
 XX
 XX Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 XX
 XX Salmonella enteritidis.
 OS
 XX WO2000060102-A2.
 PN
 XX 12-OCT-2000.
 PD
 XX 05-APR-2000; 2000WO-CA000356.
 PF
 XX 05-APR-1999; 99US-0127888P.
 PR
 XX (UYVI-) UNIV VICTORIA.
 PA
 XX White AP, Doran JL, Collison SK, Kay WW;
 PI WPI; 2000-672631/65.
 XX

DR N-PSDB; AAC64617.
 XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 FT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 135; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAf) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CagA and Agfa-homologue fimbria subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;

Query Match 89.4%; Score 693; DB 3; Length 151;
 Best Local Similarity 90.7%; Pred. No. 4.7e-59;
 Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGPDSITLSIYQYGSANAALALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGPDSITLSIYQYGSANAALALQ 60
 QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNATIDOWNAKNSDITVGYDQ 120
 DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNATIDOWNAKNSDITVGYDQ 120
 QY 121 LVTRVVTHEMAHASVMVROVGFNNATANOY 151
 DB 121 NNPAALVNQASDSSVMVROVGFNNATANOY 151

RESULT 6
 AAW23570
 ID AAW23570 standard; protein; 151 AA.
 XX
 AC AAW23570;
 XX
 XX 25-MAR-2003 (revised)
 DT 29-SEP-1997 (first entry)
 XX
 DE *Salmonella enteritidis* 27655-3b agfa.
 XX
 KW Enteropathogenic bacteria; enterobacteria; *S. enteritidis*; antibody.
 OS *Salmonella enteritidis*.
 XX
 XX Key Location/Qualifiers
 FH Misc-difference 123
 FT /note= "Encoded by GCC"
 XX

PN US5635617-A.
 XX 03-JUN-1997.
 XX 26-APR-1994; 94US-00233788.
 XX 26-APR-1993; 93US-00054452.
 XX (UVVI-) UNIV VICTORIA INNOVATION & DEV CORP.
 XX Collinson SK, Kay WW, Doran JL;
 XX WPI; 1997-309886/28.
 XX N-PSDB; AAT74142.
 PT Isolated *Salmonella* gene agfa - used for diagnosis of *Salmonella* or
 PT enteropathogenic bacteria of the Enterobacteria family.
 XX
 PS Example 2; Fig 7; 85pp; English.
 XX
 CC The present sequence represents agfa encoded by the full agfa gene
 CC derived from *Salmonella enteritidis* 27655-3b. The nucleic acid can be
 CC used to provide diagnostic assays for *Salmonella* and/or enteropathogenic
 CC bacteria of the family Enterobacteria. It can also be used to provide
 CC proteins and antibodies which can be used for assays. The nucleic acid
 CC sequence can be used to provide probes or primers which can specifically
 CC hybridise to nucleic acid molecules from greater than 99% of *Salmonella*
 CC strains that are pathogenic to warm-blooded animals relative to nucleic
 CC acid molecules from virtually all other microbial organisms. (Updated on
 CC 25-MAR-2003 to correct PF field.)
 XX
 SQ Sequence 151 AA;
 Query Match 89.3%; Score 692; DB 2; Length 151;
 Best Local Similarity 90.7%; Pred. No. 5.8e-59;
 Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGPDSITLSIYQYGSANAALALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGPDSITLSIYQYGSANAALALQ 60
 QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNATIDOWNAKNSDITVGYDQ 120
 DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNATIDOWNAKNSDITVGYDQ 120
 QY 121 LVTRVVTHEMAHASVMVROVGFNNATANOY 151
 DB 121 NNPAALVNQASDSSVMVROVGFNNATANOY 151
 RESULT 7
 AAB36355
 ID AAB36355 standard; protein; 151 AA.
 XX
 AC AAB36355;
 XX
 XX 26-FEB-2001 (first entry)
 DT
 XX
 DE Agfa::PT3#10 amino acid sequence SEQ ID NO:30.
 XX
 KW *Salmonella*; agfa; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS *Salmonella enteritidis*.
 OS *Escherichia coli*.
 XX Synthetic.
 XX
 PN WO2000060102-A2.
 XX
 XX 12-OCT-2000.
 XX
 XX 05-APR-2000; 2000WO-CA000356.
 XX

```

PR 05-APR-1999; 99US-0127888P.
XX (UYVI-) UNIV VICTORIA.
XX White AP, Doran JL, Collison SK, Kay WW;
XX WPI; 2000-672631/65.
XX N-PSDB; AAC64631.
XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX Disclosure; Page 139; 139pp; English.
XX The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbria subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX Sequence 151 AA;
SQ
Query Match 85.0%; Score 659; DB 3; Length 151;
Best Local Similarity 81.9%; Pred. No. 9.1e-56;
Matches 136; Conservative 0; Mismatches 0; Indels 30; Gaps 2;
QY 1 MKLLKVAAPAAIVVSGSALAGVVPQGGGNGHNGGNSGDPSTLSIYQGSANAALALQ 60
DB 1 MKLLKVAAPAAIVVSGSALAGVVPQGGGNGHNGGNSGDPSTLSIYQGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGQGDNSTIELTQNGFRNNATIDQWNAKNSDITVGYDQ 120
DB 61 SDARKSETTITQSGYNGADVGQGDNSTIELTQNGFRNNATIDQWNAKNSDITVGYDQ 120
QY 121 LVTRVVTHEMAHA-----SVVVRQVFGNNATANQY 151
DB 106 LVTRVVTHEMAHAGNNAALVNQTASDSVVRQVFGNNATANQY 151
RESULT 8
AAB36353
ID AAB36353 standard; protein; 151 AA.
XX AAB36353;
XX 26-FEB-2001 (first entry)
XX AgfA::P73#8 amino acid sequence SEQ ID NO:26.
XX Salmonella; agfA; chromosomal gene replacement; fimbria; epitope;
XX vaccine; immune response; immunogen.

```

```

XX Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX PN WO200060102-A2.
XX PD 12-OCT-2000.
XX PF 05-APR-2000; 2000WO-CA000356.
XX PR 05-APR-1999; 99US-0127888P.
XX (UYVI-) UNIV VICTORIA.
XX White AP, Doran JL, Collison SK, Kay WW;
XX WPI; 2000-672631/65.
XX N-PSDB; AAC64629.
XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX Disclosure; Page 138; 139pp; English.
XX The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbria subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX Sequence 151 AA;
SQ
Query Match 79.0%; Score 612; DB 3; Length 151;
Best Local Similarity 81.5%; Pred. No. 3.2e-51;
Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;
QY 1 MKLLKVAAPAAIVVSGSALAGVVPQGGGNGHNGGNSGDPSTLSIYQGSANAALALQ 60
DB 1 MKLLKVAAPAAIVVSGSALAGVVPQGGGNGHNGGNSGDPSTLSIYQGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGQGDNSTIELTQNGFRNNATIDQWNAKNSDITVGYDQ 120
DB 61 LVTRVVTHEMAHAGYNGADVGQGDNSTIELTQNGFRNNATIDQWNAKNSDITVGYDQ 120
QY 121 LVTRVVTHEMAHASVVRQVFGNNATANQY 151
DB 121 NNAALVNQTASDSVVRQVFGNNATANQY 151
RESULT 9

```

AAB36349
 ID AAB36349 standard; protein; 151 AA.
 XX
 AC AAB36349;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Agfa::PT#4 amino acid sequence SEQ ID NO:18.
 XX
 KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 PN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay WW;
 XX
 DR WPI; 2000-672631/65.
 DR N-PSDB; AAC64625.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 136; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbriin subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbriin protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;
 Query Match 78.8%; Score 611; DB 3; Length 151;
 Best Local Similarity 81.5%; Pred. No. 4e-51;
 Matches 123; Conservative 6; Mismatches 22; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVSGSALAGVVPWGGGNGGNGSSGPDSTLSITVQVGSANALALQ 60
 DB 1 MKLLKVAFAAIVSGSALAGVVPWGGGNGGNGSSGPDSTLSITVQVGSANALALQ 60

QY 61 SDARKSETTITQSGYNGGADVGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGYDQ 120
 DB 61 SDARKSETTITQSGYNGGADVGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
 QY 121 LVTRVVTHEVAHASVMVRQVFGNNTANQY 151
 DB 121 NNAALVNTQATSDSSVMVRQVFGNNTANQY 151

RESULT 10
 AAB36350
 ID AAB36350 standard; protein; 151 AA.
 XX
 AC AAB36350;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Agfa::PT#5 amino acid sequence SEQ ID NO:20.
 XX
 KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 PN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay WW;
 XX
 DR WPI; 2000-672631/65.
 DR N-PSDB; AAC64626.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 137; 139pp; English.

CC The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbriin subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbriin protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in

CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant AgfA
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 XX Sequence 151 AA;

Query Match 77.7%; Score 602; DB 3; Length 151;
 Best Local Similarity 81.5%; Pred. No. 3e-50;
 Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
 QY 61 SDARKSETTITQSGYNGADVGCGADNSTIETQNGFRNATIDQWNAKNSDITVGYDQ 120
 DB 61 SDARKSETTITQSGYNGADVGCGADNSTIETQNGFRNATIDQWNAKNSDITVGYDQ 120

RESULT 13
 AAB36348
 ID AAB36348 standard; protein; 151 AA.
 XX
 AC AAB36348;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE AgfA::PT3#3 amino acid sequence SEQ ID NO:16.
 XX
 KW Salmonella; agfA; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 FN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNTV VICTORIA.
 XX
 PI White AF, Doran JL, Collison SK, Kay WW,
 XX WPI; 2000-672631/65.
 DR

DR N-PSDB; AAC64624.
 XX
 PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant AgfA
 PT protein useful for eliciting immune response in animal.
 XX
 XX Disclosure; Page 136; 139pp; English.

CC The present invention describes a recombinant agfA gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant AgfA
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 XX Sequence 151 AA;

Query Match 74.6%; Score 578; DB 3; Length 151;
 Best Local Similarity 80.8%; Pred. No. 6.3e-48;
 Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
 QY 61 SDARKSETTITQSGYNGADVGCGADNSTIETQNGFRNATIDQWNAKNSDITVGYDQ 120
 DB 61 SDARKSETTITQSGYNGADVGCGADNSTIETQNGFRNATIDQWNAKNSDITVGYDQ 120
 QY 121 LVTRVVTHEMAHASVVMVROVGFNGNATANQY 151
 DB 121 NNAALVNQTASDSSVMVROVGFNGNATANQY 151

RESULT 14
 AAB36343
 ID AAB36343 standard; protein; 151 AA.
 XX
 AC AAB36343;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Escherichia coli CsgA amino acid sequence SEQ ID NO:7.
 XX
 KW Salmonella; agfA; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Escherichia coli.
 OS Synthetic.
 XX
 FN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX

PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
XX WPI; 2000-672631/65.
DR N-PSDB; AAC64619.
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 135; 139pp; English.
XX
XX The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (Sf17/TAf) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
XX the exemplification of the present invention
XX
SQ Sequence 151 AA;

Query Match 68.1%; Score 528; DB 3; Length 151;
Best Local Similarity 68.9%; Pred. No. 4.3e-43;
Matches 104; Conservative 20; Mismatches 27; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGSPDSTLSIYYGSAANAALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGSPDSTLSIYYGSAANAALQ 60
QY 61 SDARKSETTITQSGYNGADVGCGADNSTIETLTQNGFRNNATIDOWNAKNSDITVGYDQ 120
DB 61 TDARNSDLTITQGGGNGADVGCGSDSSIDLITQRFNGSATLDQNGKNSMTVYQFGG 120
QY 121 LVTRVVVTHEMAHSAVMVRQVGFNNATANQY 151
DB 121 GNGAAVDQTASNSVNVTVQVGFNNATAHQY 151

RESULT 15
ABR82651
ID ABR82651 standard; protein; 151 AA.
XX
AC ABR82651;
XX
DT 04-DEC-2003 (first entry)
XX
DE E. coli CsgA subunit 15 kDa protein.
XX

KW Plasma protein; immune response; antibacterial; vaccine; gene therapy.
XX
XX Escherichia coli.
XX
XX WO2003064446-A2.
XX
XX 07-AUG-2003.
XX
XX 30-JAN-2003; 2003WO-BP000943.
XX
XX 31-JAN-2002; 2002GB-00002275.
XX
XX (HANS-) HANSA MEDICAL RES AB.
XX
XX Bjoerck L, Olsen A, Wikstroem M, Herwald H;
XX
XX WPI; 2003-646136/61.
DR N-PSDB; ACF36153.
XX
XX New isolated peptide capable of binding a mammalian plasma protein,
PT useful in the manufacture of a medicament for the prevention and/or
PT treatment of a bacterial infection, such as Escherichia coli, Salmonella
PT or Shigella infections.
XX
XX Disclosure; Page 41-42; 42pp; English.
XX
XX The invention relates to an isolated peptide capable of binding a
CC mammalian plasma protein or of generating an immune response in a mammal
CC selected from sequences shown in ABR82642, ABR82648-49. The peptide or
CC antibody is useful for treating a bacterial infection in a human or
CC animal or in the manufacture of a medicament for the prophylactic
CC treatment of a bacterial infection, such as Escherichia coli, Salmonella
CC or Shigella infection. The peptide that is immobilized on a solid support
CC is also useful as a reagent for determining the ability of a plasma
CC protein to bind to bacteria. The present sequence represents an E. coli
CC 15 kDa protein
XX
SQ Sequence 151 AA;
Query Match 67.5%; Score 523; DB 7; Length 151;
Best Local Similarity 68.2%; Pred. No. 1.3e-42;
Matches 103; Conservative 20; Mismatches 28; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGSPDSTLSIYYGSAANAALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGSPDSTLSIYYGSAANAALQ 60
QY 61 SDARKSETTITQSGYNGADVGCGADNSTIETLTQNGFRNNATIDOWNAKNSDITVGYDQ 120
DB 61 TDARNSDLTITQGGGNGADVGCGSDSSIDLITQRFNGSATLDQNGKNSMTVYQFGG 120
QY 121 LVTRVVVTHEMAHSAVMVRQVGFNNATANQY 151
DB 121 GNGAAVDQTASNSVNVTVQVGFNNATAHQY 151
Search completed: August 2, 2004, 14:48:24
Job time : 44.9 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:40:48 ; Search time 12 Seconds
(without alignments)
649.627 Million cell updates/sec

Title: US-09-543-407-14
 Perfect score: 775
 Sequence: 1 MKLLKVAEFAAIVVSSALA.....HASVTVRQVGFQNNATANQY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 secs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

```
Database : Issued_Patents_AA.*
1: /cgm2_6/ptodata/2/iaa/5A COMB rep.*
2: /cgm2_6/ptodata/2/iaa/5B COMB rep.*
3: /cgm2_6/ptodata/2/iaa/6A COMB rep.*
4: /cgm2_6/ptodata/2/iaa/6B COMB rep.*
5: /cgm2_6/ptodata/2/iaa/PTUS COMB rep.*
6: /cgm2_6/ptodata/2/iaa/backfills1 rep.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	692	89.3	151	1	US-08-233-788A-59	Sequence 59, Appl
2	514	66.3	120	1	US-08-233-788A-57	Sequence 57, Appl
3	92	11.9	673	3	US-09-196-387-8	Sequence 8, Appl
4	92	11.9	673	4	US-09-841-835-8	Sequence 8, Appl
5	92	11.9	949	3	US-09-196-387-10	Sequence 10, Appl
6	92	11.9	949	4	US-09-841-835-10	Sequence 10, Appl
7	92	11.9	1327	3	US-09-196-387-2	Sequence 2, Appl
8	92	11.9	1327	4	US-09-841-835-2	Sequence 2, Appl
9	92	11.9	1327	4	US-09-972-115A-8	Sequence 8, Appl
10	89.5	11.5	738	3	US-08-864-038A-3	Sequence 3, Appl
11	87	11.2	2123	3	US-08-968-685A-10	Sequence 10, Appl
12	83.5	10.8	339	4	US-09-252-991A-32096	Sequence 32096, A
13	83.5	10.8	943	4	US-09-056-556-204	Sequence 204, App
14	83.5	10.8	943	4	US-09-072-596-139	Sequence 139, App
15	83.5	10.8	943	4	US-09-477-135A-131	Sequence 131, App
16	83.5	10.8	943	4	US-09-072-967-204	Sequence 204, App
17	82.5	10.6	2736	4	US-09-252-991A-30227	Sequence 30227, App
18	81	10.5	878	4	US-09-540-238-3401	Sequence 3401, App
19	80.5	10.4	873	4	US-09-336-447A-13	Sequence 13, Appl
20	80	10.3	892	4	US-09-336-447A-5	Sequence 5, Appl
21	79.5	10.3	204	6	5187262-1	Patent No. 5187362
22	79.5	10.3	518	3	US-09-043-123-2	Sequence 2, Appl
23	79.5	10.3	1007	4	US-09-489-039A-11518	Sequence 11518, A
24	78.5	10.1	539	4	US-09-719-402A-2	Sequence 2, Appl
25	78.5	10.1	714	4	US-09-841-786-4	Sequence 4, Appl
26	78.5	10.1	1912	1	US-08-409-995-4	Sequence 4, Appl
27	78.5	10.1	1912	3	US-08-685-467-4	Sequence 4, Appl

ALIGNMENTS

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

Query Match 89.3%; Score 692; DB 1; Length 151;
Best Local Similarity 90.7%; Pred. No. 3.9e-62;
Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

Qy 1 MKLLKVAFAAIIVVSGSALGVVPPQWGGGNNHGGNSGDPSTLSIYVYGSANAAALQ 60

```

Db      1  MKLLKVAFAAIVVSSALAGVVPQWGGGNGHNGSGGPDSTLSIYQGSANAALALQ 60
QY      61  SPARKSETTITQSGYNGADVCGGADNSTIETLQNGFRNNATIDOWNAKNSDITVQGYDQ 120
Db      61  SPARKSETTITQSGYNGADVCGGADNSTIETLQNGFRNNATIDOWNAKNSDITVQGYGG 120
QY      121  LVTRVVTHEMAHSAVMVRQVGFNNATANOY 151
Db      121  NNPALVNTQASDSSVMVRQVGFNNATANOY 151

RESULT 2
US-08-233-788A-57
; Sequence 57, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, Karen S.
; APPLICANT: Clouthier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,788A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 120 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-233-788A-57

Query Match 66.3%; Score 514; DB 1; Length 120;
Best Local Similarity 100.0%; Pred. No. 2.2e-44;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      22  VVPQGGGNGHNGSGGPDSTLSIYQGSANAALALQSDARKSETTITQSGYNGADY 81
Db      1  VVPQGGGNGHNGSGGPDSTLSIYQGSANAALALQSDARKSETTITQSGYNGADY 60
QY      82  GQADNSTIETLQNGFRNNATIDOWNAKNSDITVQGY 118
Db      61  GQADNSTIETLQNGFRNNATIDOWNAKNSDITVQGY 97

RESULT 3
US-09-196-387-8
; Sequence 8, Application US/09196387
; Patent No. 6277613
; GENERAL INFORMATION:

```

```

; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; TITLE OF INVENTION: OF USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/196,387
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/095,225
; FILING DATE: June 10, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 673 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-196-387-8

Query Match 11.9%; Score 92; DB 3; Length 673;
Best Local Similarity 30.4%; Pred. No. 0.54;
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

QY      6  VAAFAAL-VVGSALAGVVPQWGGGNGHNGSGGPDSTLSIYQGSANAALALQSDAR 64
Db      99  VAAAPVVPVAVSTSSAAGVAPNPAGSGNNPSSSSPTSS-SSSSPSSPGSLAESPEAA 157
QY      65  KSETTIT---QSGYNGADVCGGADNSTIETLQNG--FRNNATIDOWNAKNSDI 113
Db      158  GVSSTAPLPGGAAGPGTGVPAVSGALRELLEACRNGDVSRVRLVDAANVNAKM 212

RESULT 4
US-09-841-835-8
; Sequence 8, Application US/09841835
; Patent No. 6506587
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; TITLE OF INVENTION: OF USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 949 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-196-387-10

Query Match 11.9%; Score 92; DB 3; Length 949;
Best Local Similarity 30.4%; Pred. No. 0.84;
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

QY 6 VAAPAAI-VVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIVQYGSANAALQSDAR 64
DB 99 VAAAPVVPVSTSSAAGVAPNPAGSGNNNSPPSSSPTSS-SSSSPSPGSSIAESPEA 157
QY 65 KSETTIT---QSGYNGADVGGAGDNSTIELTQNG--FRNNATIDOWNAKNSDI 113
DB 158 GYVSTAPLGFGAAGPTGVPVAVSGALRELEACENGVDVSKVLVDAAVNAKDM 212

RESULT 6
US-09-841-835-10
; Sequence 10, Application US/09841835
; Patent No. 6506587
; GENERAL INFORMATION:
; APPLICANT: de Large, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; TITLE OF INVENTION: OF USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Johnson
; STREET: 411 Hackensack Avenue, 4th Floor

```

```

? COUNTRY: USA
? ZIP: 07601
?
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/09/841,835
? FILING DATE:
? CLASSIFICATION:
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 09/196,387
? FILING DATE:
? ATTORNEY/AGENT INFORMATION:
? NAME: Jackson Esq., David A.
? REGISTRATION NUMBER: 26,742
? REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 201-487-5800
? TELEFAX: 201-343-1684
? TELEX: 133521
? INFORMATION FOR SEQ ID NO: 10:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 949 amino acids
? TYPE: amino acid
? TOPOLOGY: linear
? MOLECULE TYPE: protein
? US-09-841-835-10
?
? Query Match 11.9%; Score 92; DB 4; Length 949;
? Best Local Similarity 30.4%; Pred.No. 0.84;
? Matches 35; Conservative 15; Mismatches 57; Incels 8; Gaps 4
?
? 6 VAAFAAI-VVSGSALAGVFPQWGGGNNHGGGNSGPDSTLSIVQYSANLALQSDAR 64

```

Db 99 VAAAPVPAVSTSSAAGVAPNPAAGSGNNSSPSSSPTSS-SSSPSPSGSLAESPEAA 157
QY 65 KSETTIT---OSGYNGADVGQADNSTIETQNG--FRNNATIDOWNAKNSDI 113
Db 158 GVSSTAPLPGGAAGPCTGVPVAVSGALRELLACRNGDVSVRKRLVDAANVNAKDM 212

RESULT 7

US-09-196-387-2
; Sequence 2, Application US/09196387
; Patent No. 6277613
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; TITLE OF INVENTION: OF USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/196,387
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/095,225
; FILING DATE: June 10, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1327 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO

US-09-196-387-2

Query Match 11.9%; Score 92; DB 3; Length 1327;
Best Local Similarity 30.4%; Pred. No. 1.3;
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;
QY 6 VAAFAAI-VVGSALAGVVPQGGGNNHNGGNSGPDSTLSIYQVGSANAALQSDAR 64
Db 99 VAAAPVPAVSTSSAAGVAPNPAAGSGNNSSPSSSPTSS-SSSPSPSGSLAESPEAA 157
QY 65 KSETTIT---OSGYNGADVGQADNSTIETQNG--FRNNATIDOWNAKNSDI 113
Db 158 GVSSTAPLPGGAAGPCTGVPVAVSGALRELLACRNGDVSVRKRLVDAANVNAKDM 212

RESULT 8

US-09-841-835-2
; Sequence 2, Application US/09841835
; Patent No. 6506587
; GENERAL INFORMATION:

; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; TITLE OF INVENTION: OF USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/841,835
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/196,387
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1327 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO

US-09-841-835-2

Query Match 11.9%; Score 92; DB 4; Length 1327;
Best Local Similarity 30.4%; Pred. No. 1.3;
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;
QY 6 VAAFAAI-VVGSALAGVVPQGGGNNHNGGNSGPDSTLSIYQVGSANAALQSDAR 64
Db 99 VAAAPVPAVSTSSAAGVAPNPAAGSGNNSSPSSSPTSS-SSSPSPSGSLAESPEAA 157
QY 65 KSETTIT---OSGYNGADVGQADNSTIETQNG--FRNNATIDOWNAKNSDI 113
Db 158 GVSSTAPLPGGAAGPCTGVPVAVSGALRELLACRNGDVSVRKRLVDAANVNAKDM 212

RESULT 9

US-09-972-115A-8
; Sequence 8, Application US/09972115A
; Patent No. 6599728
; GENERAL INFORMATION:
; APPLICANT: Geron Corporation
; APPLICANT: Gregg, Morin B.
; APPLICANT: Walter, Funk D.
; APPLICANT: Mieczyslaw, Piatyszek A.
; TITLE OF INVENTION: A Second Mammalian Telomerase
; FILE REFERENCE: 080/003C
; CURRENT APPLICATION NUMBER: US/09/972,115A
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: US 60/128,577
; PRIOR FILING DATE: 2000-04-10
; PRIOR APPLICATION NUMBER: US 60/129,123
; PRIOR FILING DATE: 1999-04-13
; NUMBER OF SEQ ID NOS: 64


```

; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 1327
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-972-115A-8

Query Match          11.2%; Score 92; DB 4; Length 1327;
Best Local Similarity 30.4%; Pred. No. 1.3;
Matches    35; Conservative 15; Mismatches   57; Indels      8; Gaps     4;

QY      6 VAAFAAI-VVSGSALAGVVFPQWGGGNHNGGNSGPDSTLSIYQYGSAALAQSDAR 64
       ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db      99 VAAAAPVPAVTS::SSAAGVAFNPAGSGNNSPSSSSPTSS-SSSSPSFGSLAESPEAA 157
       ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :

QY      65 KSEHTIT----OSGVNGADVCQGADNSTIELTQNG--FRNNAIDQWNKNSDI 113
       ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db      158 GVSSTAPLGGAAGPGTGVPVSGALRELLEACRNGDYSRVKRLVDYAANNAKOM 212
       ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :

RESULT 10
US-08-864-038A-3
; Sequence 3, Application US/08864038A
; Patent No. 6001592
; GENERAL INFORMATION:
; APPLICANT: Kunio NAKASHIMA et al.
; TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE cDNA, VECTOR
; TITLE OF INVENTION: CONTAINING SAID cDNA, HOST CELLS TRANSFORMED WITH SAID
; TITLE OF INVENTION: VECTOR. POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING
; TITLE OF INVENTION: SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBODY
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: 812-5 Hirano
; STREET: Isshinden
; CITY: Tsu-city
; STATE: Mie-prefecture
; COUNTRY: JAPAN
; ZIP: 514-01
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Microsoft Windows 95
; SOFTWARE: Word Perfect 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/864,038A
; FILING DATE: May 28, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-184459
; FILING DATE: 15-July-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: C. Bruce Hamburg
; REGISTRATION NUMBER: 22,389
; REFERENCE/DOCKET NUMBER: F-5610
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)986-2340
; TELEFAX: (212)953-7733
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 738
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Pinctada fucata
; CELL TYPE: mantle epithelial cell
; FEATURE:
; NAME/KEY: peptide
; LOCATION: from 1 to 738
; IDENTIFICATION METHOD: E (by experiment)
;
US-08-864-038A-3

Query Match          11.5%; Score 89.5; DB 3; Length 738;

```

Db 213 NAFGRATA 221

RESULT 12

US-09-252-991A-32096

Sequence 32096 Application US/09252991A

Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A

CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 32096

LENGTH: 339

TYPE: PRT

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-32096

Query Match 10.8%; Score 83.5; DB 4; Length 339;

Best Local Similarity 21.5%; Pred. No. 1.6;

Matches 35; Conservative 28; Mismatches 53; Indels 47; Gaps 7;

QY 5 KVAAPAAIVVSGALAGVVPQWG-----CGNNGGNSGSPDSTLSIYQYGSANAAL 57

Db 58 KVSNFCTL--NNAVSQSIKIDAGNVGVNVAAGDNNQQAALASADASFVGTATAS- 114

QY 58 ALQSDARKSETITQSGYNGADVGQADNSIETLTQNGFRNNATIDQWNAKNS----- 111

Db 115 -----TSVLQGYGN-----TLNYSNPNTASLSNANVSGNLG 149

QY 112 -DITVGYDQ-----LVTRVVTHEMAHASVWVQVGFNNATAN 149

Db 150 VNVAGNFNQKNDLAAAVSNQYSTAGSAAQTSTG-NTTVN 191

RESULT 13

US-09-056-556-204

Sequence 204 Application US/09056556

Patent No. 6350456

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.

APPLICANT: Skeiky, Yasir A.W.

APPLICANT: Dillon, David C.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND

NUMBER OF SEQUENCES: 241

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/056,556

FILING DATE: 07-APR-1998

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Maki, David J.

REGISTRATION NUMBER: 31,392

REFERENCE/DOCKET NUMBER: 210121.457

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 622-4900

INFORMATION FOR SEQ ID NO: 199:

SEQUENCE CHARACTERISTICS:

LENGTH: 943 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

US-09-056-556-204

Query Match 10.8%; Score 83.5; DB 4; Length 943;

Best Local Similarity 25.4%; Pred. No. 6;

Matches 29; Conservative 16; Mismatches 50; Indels 19; Gaps 4;

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 204:

SEQUENCE CHARACTERISTICS:

LENGTH: 943 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

US-09-056-556-204

Query Match 10.8%; Score 83.5; DB 4; Length 943;

Best Local Similarity 25.4%; Pred. No. 6;

Matches 29; Conservative 16; Mismatches 50; Indels 19; Gaps 4;

QY 12 IVVSGSALAGVVPQWGCGNHNGGNSGSPDSTLSIYQYGSANAALALQSDAR---KSET 68

Db 159 IGTGSGLLGF-----GGLNSGTGN-----IGLNSGTGTVGNSGTGNGWIGNSG 205

QY 69 TITQSGYNGADVGQADNSIETLTQNGFRNNATIDQWNAKNSD---ITVGYD 119

Db 206 NSYNTGFGNSGDANTGFNSGIANTGVNAGNVTGNSYNGPNSNTGCFNMGQYN 259

RESULT 14

US-09-072-596-199

Sequence 199 Application US/09072596

Patent No. 6458366

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.

APPLICANT: Skeiky, Yasir A.W.

APPLICANT: Dillon, David C.

APPLICANT: Campos-Neto, Antonia

APPLICANT: Houghton, Raymond

APPLICANT: Vedvick, Thomas S.

APPLICANT: Twardzik, Daniel R.

APPLICANT: Lodes, Michael J.

APPLICANT: Hendrickson, Ronald C.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF

NUMBER OF SEQUENCES: 350

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/072,596

FILING DATE: 05-MAY-1998

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Maki, David J.

REGISTRATION NUMBER: 31,392

REFERENCE/DOCKET NUMBER: 210121.417C9

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 199:

SEQUENCE CHARACTERISTICS:

LENGTH: 943 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

US-09-072-596-199

Query Match 10.8%; Score 83.5; DB 4; Length 943;

Best Local Similarity 25.4%; Pred. No. 6;

Matches 29; Conservative 16; Mismatches 50; Indels 19; Gaps 4;

```
QY 12 IIVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQYGSANAALALQSDAR---KSET 68
Db 159 IGLTGSGLGF-----GGLNSGTGN-----IGLNSGTGNVGVGNSGTGNWGICNSG 205
QY 69 TITOSGYGNGADVCGGADNSTIELTQNGFRNNATIDOWNAKNSD---ITVGOYD 119
Db 206 NSYNTGFGNSGDANTGFFNSGIANTGVGNAGNYNTGTSYNPNSNTGGFNMGCYN 259
```

RESULT 15

```
US-09-477-135A-131
; Sequence 131, Application US/09477135A
; Patent No. 6572865
; GENERAL INFORMATION:
; APPLICANT: Nano, Francis
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
; TITLE OF INVENTION: immunostimulatory Peptides
; FILE REFERENCE: 52888
; CURRENT APPLICATION NUMBER: US/09/477,135A
; CURRENT FILING DATE: 2000-01-03
; PRIOR APPLICATION NUMBER: 08990823
; PRIOR FILING DATE: 1997-12-15
; PRIOR APPLICATION NUMBER: US 96/10375
; PRIOR FILING DATE: 1996-06-14
; PRIOR APPLICATION NUMBER: 60/000,254
; PRIOR FILING DATE: 1995-06-15
; NUMBER OF SEQ ID NOS: 169
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 131
; LENGTH: 943
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-477-135A-131
```

```
Query Match 10.8%; Score 83.5; DB 4; Length 943;
Best Local Similarity 25.4%; Pred. No. 6;
Matches 29; Conservative 16; Mismatches 50; Indels 19; Gaps 4;

QY 12 IIVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQYGSANAALALQSDAR---KSET 68
Db 266 IGLTGSGLGF-----GGLNSGTGN-----IGLNSGTGNVGVGNSGTGNWGICNSG 312
QY 69 TITOSGYGNGADVCGGADNSTIELTQNGFRNNATIDOWNAKNSD---ITVGOYD 119
Db 313 NSYNTGFGNSGDANTGFFNSGIANTGVGNAGNYNTGTSYNPNSNTGGFNMGCYN 366
```

Search completed: August 2, 2004, 14:58:32
Job time : 13 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:54:48 ; Search time 36.8 Seconds
(without alignments)
1287.123 Million cell updates/sec

Title: US-09-543-407-14

Perfect score: 775

Sequence: 1 MKLLKVAFAAIVVSGSALA.....HASVMVRQVFGNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	525	67.7	151	12	US-09-741-873B-4
2	525	67.7	151	12	US-09-741-873B-4
3	447	57.7	131	12	US-09-741-873B-2
4	447	57.7	131	12	US-09-741-873B-2
5	104	13.4	445	15	US-10-369-493-20638
6	98	12.6	210	12	US-10-425-114-53421
7	98	12.6	211	12	US-10-425-599-245046
8	95	12.3	186	12	US-10-282-122A-49412
9	92	11.9	673	9	US-09-841-835-8
10	92	11.9	949	9	US-09-841-835-10
11	92	11.9	1327	9	US-09-841-835-2
12	92	11.9	1327	10	US-09-972-115A-8
13	92	11.9	1327	14	US-10-199-937-4
14	90.5	11.7	263	12	US-10-425-114-49960
15	88.5	11.4	145	16	US-10-437-963-147748

16	88	11.4	369	12	US-10-425-114-56041	Sequence 56041, A
17	88	11.4	486	12	US-10-424-599-275468	Sequence 275468, A
18	88	11.4	507	12	US-10-425-114-57763	Sequence 57763, A
19	87	11.2	597	9	US-09-793-306-146	Sequence 146, App
20	87	11.2	628	12	US-10-282-122A-53269	Sequence 53269, A
21	87	11.2	678	12	US-10-282-122A-64573	Sequence 64573, A
22	87	11.2	2122	9	US-09-813-214A-9	Sequence 9, Appli
23	86.5	11.2	527	9	US-09-712-363-156	Sequence 156, App
24	86.5	11.2	527	14	US-10-080-170-348	Sequence 348, App
25	86.5	11.2	527	16	US-10-080-170-348	Sequence 348, App
26	86	11.1	482	14	US-10-156-761-8763	Sequence 8763, App
27	86	11.1	486	15	US-10-369-493-20619	Sequence 20619, A
28	85.5	11.0	270	16	US-10-437-963-122263	Sequence 122263, A
29	85.5	11.0	1048	14	US-10-174-383-56	Sequence 56, Appl
30	85.5	11.0	1048	15	US-10-374-780A-2086	Sequence 2086, App
31	85	11.0	438	14	US-10-156-761-9343	Sequence 9343, App
32	84.5	10.9	271	14	US-10-156-761-11721	Sequence 11721, A
33	83.5	10.8	943	9	US-09-996-634-131	Sequence 131, App
34	83.5	10.8	943	10	US-09-997-182-131	Sequence 131, App
35	83.5	10.8	943	10	US-09-997-181-131	Sequence 199, App
36	83.5	10.8	943	14	US-10-193-002-199	Sequence 204, App
37	83.5	10.8	943	14	US-10-084-843-204	Sequence 64369, A
38	83.5	10.8	3300	12	US-10-282-122A-64369	Sequence 61520, A
39	83	10.7	242	12	US-10-425-114-61520	Sequence 1136, App
40	83	10.7	254	10	US-09-880-748-1136	Sequence 1136, App
41	83	10.7	254	10	US-09-880-748-1136	Sequence 1136, App
42	83	10.7	254	12	US-10-293-418-1136	Sequence 1136, App
43	83	10.7	254	12	US-10-293-418-1136	Sequence 1165, App
44	83	10.7	257	10	US-09-880-748-1494	Sequence 1494, App
45	83	10.7	257	12	US-10-293-418-1494	Sequence 1494, App

ALIGNMENTS

RESULT 1

US-09-741-873B-4
; Sequence 4, Application US/09741873B
; Publication No. US20020081722A1
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-4

Query Match 67.7%; Score 525; DB 12; Length 151;

Best Local Similarity 68.2%; Pred. No. 2.4e-45;

Matches 103; Conservative 21; Mismatches 27; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVWPQWGGGNGHNGGNSGPGDTLSIYQYGSANAALQ 60

```
Db 1 MKLLKVAIAAIVFSSAGVVPQYGGGNGHGGGNSGPNSELNIYQYGGNSALALQ 60
Qy 61 SDARKSETTTTQSGYNGADVGQADNSTIELTQNGFRNATIDQWAKNSDITVQGYDQ 120
Db 61 TDARNSDLTITQHGNGADVGQSDSSIDLTLQRFSGNSATLDQWNGKNSMTVKQFGG 120
Qy 121 LVTRVVTHEMAHASVMVRQVGFNNATANOY 151
Db 121 GNGAAVDQTASNSVNVTVQGFNNATAHOY 151
```

RESULT 2

```
US-09-741-873B-4
; Sequence 4, Application US/09741873B
; Publication No. US20040096965A9
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR FILING DATE: 1997-11-26
; PRIOR FILING DATE: 1997-11-26
; PRIOR FILING DATE: 1989-05-04
; PRIOR FILING DATE: 1989-05-04
; PRIOR FILING DATE: 1991-11-06
; PRIOR FILING DATE: 1992-11-03
; PRIOR FILING DATE: 1994-01-28
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
```

US-09-741-873B-4

```
Query Match 67.7%; Score 525; DB 12; Length 151;
Best Local Similarity 68.2%; Pred. No. 2.1e-45;
Matches 103; Conservative 21; Mismatches 27; Indels 0; Gaps 0;

Qy 1 MKLLKVAIAAIVFSSAGVVPQYGGGNGHGGGNSGPNSELNIYQYGSANAALALQ 60
Db 1 MKLLKVAIAAIVFSSAGVVPQYGGGNGHGGGNSGPNSELNIYQYGGNSALALQ 60
Qy 61 SDARKSETTTQSGYNGADVGQADNSTIELTQNGFRNATIDQWAKNSDITVQGYDQ 120
Db 61 TDARNSDLTITQHGNGADVGQSDSSIDLTLQRFSGNSATLDQWNGKNSMTVKQFGG 120
Qy 121 LVTRVVTHEMAHASVMVRQVGFNNATANOY 151
Db 121 GNGAAVDQTASNSVNVTVQGFNNATAHOY 151
```

RESULT 3

```
US-09-741-873B-2
; Sequence 2, Application US/09741873B
; Publication No. US20020081722A1
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; CURRENT FILING DATE: 2003-04-04
```

```
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 2
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-2
```

Query Match 57.7%; Score 447; DB 12; Length 131;

Best Local Similarity 64.9%; Pred. No. 1.5e-37;

Matches 85; Conservative 20; Mismatches 26; Indels 0; Gaps 0;

```
Qy 21 GVFPQGGGNGHGGGNSGPNSELNIYQYGSANAALALQSDARKSETTTTQSGYNGAD 80
Db 1 GVFPQGGGNGHGGGNSGPNSELNIYQYGGNSALALQTDARNSDLTITQHGNGAD 60
Qy 81 VQGAADNSTIELTQNGFRNATIDQWAKNSDITVQGYDOLVTRVTHEMAHASVMVRQV 140
Db 61 VQGSDDSSIDLTLQRFSGNSATLDQWNGKNSMTVKQFGGNGAAYDQTASNSVNVTVQ 120
Qy 141 GFGNNATANOY 151
Db 121 GFGNNATAHOY 131
```

RESULT 4

```
US-09-741-873B-2
; Sequence 2, Application US/09741873B
; Publication No. US20040096965A9
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; CURRENT FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 2
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-2
```

Query Match 57.7%; Score 447; DB 12; Length 131;

```
Best Local Similarity 64.9%, Pred. No. 1.5e-37;
Matches 85; Conservative 20; Mismatches 26; Indels 0; Gaps 0;

QY 21 GVPQWGGGNGHNGSGSPDSTLSIYQGSANAALALQSDARKSETTITQSGYNGAD 80
D 1 GVPQYGGGNGHGGGNGSGSELTNYQYGGNSALALQTDARNSDLTITQHGSGNGAD 60
QY 81 VQGGADNSTIELTQNGFRNNATIDOWNNAKNSDITVQYDQDLVTRVTHEMAHASVMYQV 140
D 61 VQGGSDSSIDLTORFGNSATLDQWNGKNSMTVKQGGNGGAADVQDTASNSSVNTQV 120
QY 141 GFGNNATANQY 151
D 121 GFGNNATAHQY 131

RESULT 5
US-10-369-493-20638
; Sequence 20638, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 20638
; LENGTH: 445
; TYPE: PRT
; ORGANISM: Rhodospseudomonas palustris
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(445)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-20638

Query Match 13.4%; Score 104; DB 15; Length 445;
Best Local Similarity 26.6%; Pred. No. 0.051;
Matches 45; Conservative 25; Mismatches 61; Indels 38; Gaps 8;

QY 7 AAFAA-----TWGSGALAGVVPQWGGG-----NHGG-----GNSGGPDSTLSIYQY 50
D 19 AAFADSNVTYLNQTGNDQANITQSGNSYGAFGNSGFLQENGTLTSGA-NLLTVKQS 77
QY 51 GSANAALALQSDARKSETTITQSGYNGADVQGGADNSTIELTQNGFRNNATIDOWNAKN 110
D 78 GNSNV-----GRDIOGKSGAGNSAIFQETGSDVELQQTGTSNGAVPSGWNVTN 129
QY 111 SDITVQYDQDLVTRVTHEMAHASVM-----VRQVGFNNATANQ 150
D 130 DP---GVFNK-ITQSSSSNGSKSVITQDGKNVFSIKQGTGNTGNTSVNQ 174

RESULT 6
US-10-425-114-53421
; Sequence 53421, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
```

```
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 53421
; LENGTH: 210
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: 700839445_FLI.pep
US-10-425-114-53421
```

Query Match 12.6%; Score 98; DB 12; Length 210;

Best Local Similarity 26.7%; Pred. No. 0.08; Mismatches 16; Gaps 4;

Matches 36; Conservative 18; Indels 65; Indels 16; Gaps 4;

QY 26 WGGGNGHNGGNSGSPDSTLSIYQ-----GSANAALALQSDARKSETTITQSGYNG 78

D 61 WGGGGGGGGGGESASNVATYHYEPEQHGWDLNAVSAYCSTWDASKPYSRKYGT 120

QY 79 ADVG---CGADN--STIELTQNGFRNNA--TIDOWNNAKNSDITVQYDQDLVTRVTHE 129

D 121 AFCGPVGPGRDSCGKCLRVNTGTGANTIVRIVDOCSNGGLDLDVGVFNRIDTGRGYQ 180

QY 130 MAHASVMVRQVGFN 144

D 181 QGHLIVNYQFVDCGN 195

RESULT 7

US-10-424-599-245046

; Sequence 245046, Application US/10424599

; Publication No. US20040031072A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa Thomas J

; APPLICANT: Kovalic David K

; APPLICANT: Zhou Yihua

; APPLICANT: Cao Yongwei

; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With

; FILE REFERENCE: 38-21(53223)B

; CURRENT APPLICATION NUMBER: US/10/424,599

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 285684

; SEQ ID NO 245046

; LENGTH: 211

; TYPE: PRT

; ORGANISM: Glycine max

; FEATURE:

; OTHER INFORMATION: Clone ID: PAT_MRT3847_63306C.1.pep

US-10-424-599-245046

Query Match 12.6%; Score 98; DB 12; Length 211;

Best Local Similarity 26.7%; Pred. No. 0.08;

Matches 36; Conservative 18; Mismatches 65; Indels 16; Gaps 4;

QY 26 WGGGNGHNGGNSGSPDSTLSIYQ-----GSANAALALQSDARKSETTITQSGYNG 78

D 62 WGGGGGGGGGGESASNVATYHYEPEQHGWDLNAVSAYCSTWDASKPYSRKYGT 121

QY 79 ADVG---CGADN--STIELTQNGFRNNA--TIDOWNNAKNSDITVQYDQDLVTRVTHE 129

D 122 AFCGPVGPGRDSCGKCLRVNTGTGANTIVRIVDOCSNGGLDLDVGVFNRIDTGRGYQ 181

QY 130 MAHASVMVRQVGFN 144

D 182 QGHLIVNYQFVDCGN 196

RESULT 8

US-10-282-122A-49412

; Sequence 49412, Application US/10282122A

```
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Orlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 49412
; LENGTH: 186
; TYPE: PRI
; ORGANISM: Burkholderia fungorum
; US-10-282-122A-49412

Query Match      12.3%; Score 95; DB 12; Length 186;
Best Local Similarity 29.9%; Pred. No. 0.14;
Matches 50; Conservative 20; Mismatches 59; Indels 38; Gaps 11;

QY 1 MKLLKVAAPAAIVVSGSALAGVWPQGGNGNNGGNS- - - - -GPDSTLSIYQVGSANAA 56
   ||| : : : : : |||
Db 1 MKL-----SLPAIAAASSFLGLVPL- - - - -GAHAADGTISITGTVTDTSCSI--NGNANG- 49
   ||| : : : : : |||

QY 57 LALQSDARKSETITQSGVNGADVG--QGADNST--IELTQNG- - - - -FRNNAT 102
   ||| : : : : : |||
Db 50 ----TPADKATLATVPA--GSLASAGAVAGTGNPTDLQLSLTGTCTGTATKAIARFENGPT 104
   ||| : : : : : |||

QY 103 IDQWNAKNSDITVGQYDQVATRVVTHMAHASVMVQVGFNGNATAN 149
   ||| : : : : : |||
Db 105 VDQTGYLSN-TAGTAQRVVELLNAQMOPINV---TTGANNDIITN 147
   ||| : : : : : |||

RESULT 9
US-09-841-835-8
; Sequence 8, Application US/09841835
; Patent No. US20020076795A1
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; TITLE OF INVENTION: OF USE THEREOF

; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/841,835
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/196,387
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 673 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-841-835-8

Query Match      11.9%; Score 92; DB 9; Length 673;
Best Local Similarity 30.4%; Pred. No. 1.4;
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

QY 6 VAAFAAI-VVSGSALAGVWPQGGNGNNGGNSGPDSTLSIYQVGSANAAALQSDAR 64
   ||| : : : : : |||
Db 99 VAAAPVPAVSTSSAAGVAPNPAGSGNSNPSSSSSTSS-SSSSPSSPGSSLAESPRA 157
   ||| : : : : : |||

QY 65 KSETTIT- - - - -QSGYNGADVGQGDADNSTIELTQNG- - - - -FRNNATIDQWNAKNSDI 113
   ||| : : : : : |||
Db 158 GVSSTAPLGPAGPGTGPVAVSGALRELLEACRNGDVSRRKLVDAANVNAKDM 212
   ||| : : : : : |||

RESULT 10
US-09-841-835-10
; Sequence 10, Application US/09841835
; Patent No. US20020076795A1
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; TITLE OF INVENTION: OF USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/841,835
```



```

; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/196,387
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 949 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-841-835-10

Query Match 11.9%; Score 92; DB 9; Length 949;
Best Local Similarity 30.4%; Pred. No. 2.2;
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

Qy 6 VAAPAAI-VVGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQYGSANAALALQSDAR 64
Db 99 VAAAPVPAVSTSSAAGVAPNPAGSGNNSPSSSSPTSS-SSSSPSPGSSLAESPEAA 157

Qy 65 KSETTIT---QSGYNGADVGQGDNSTIELTQNG--FRNNATIDQWNAKNSDI 113
Db 158 GVSSTAPLPGAAGPGTGVPAVSGALRELLEACRNGDVSRVKRLVDAANVNAKDM 212

RESULT 11
US-09-841-835-2
; Sequence 2, Application US/09841835
; Patent No. US20020076795A1
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/841,835
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/196,387
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1327 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; US-09-841-835-2

Query Match 11.9%; Score 92; DB 9; Length 1327;
Best Local Similarity 30.4%; Pred. No. 3.3;
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

Qy 6 VAAPAAI-VVGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQYGSANAALALQSDAR 64
Db 99 VAAAPVPAVSTSSAAGVAPNPAGSGNNSPSSSSPTSS-SSSSPSPGSSLAESPEAA 157

Qy 65 KSETTIT---QSGYNGADVGQGDNSTIELTQNG--FRNNATIDQWNAKNSDI 113
Db 158 GVSSTAPLPGAAGPGTGVPAVSGALRELLEACRNGDVSRVKRLVDAANVNAKDM 212

RESULT 12
US-09-972-115A-8
; Sequence 8; Application US/09972115A
; Publication No. US20030032769A1
; GENERAL INFORMATION:
; APPLICANT: Geron Corporation
; APPLICANT: Gregg, Morin B.
; APPLICANT: Walter, Funk D.
; APPLICANT: Mieczyslaw, Piatyszek A.
; TITLE OF INVENTION: A Second Mammalian Telomerase
; FILE REFERENCE: 080/003C
; CURRENT APPLICATION NUMBER: US/09/972,115A
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: US 60/128,577
; PRIOR FILING DATE: 2000-04-10
; PRIOR APPLICATION NUMBER: US 60/129,123
; PRIOR FILING DATE: 1999-04-13
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 8
; LENGTH: 1327
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-972-115A-8

Query Match 11.9%; Score 92; DB 10; Length 1327;
Best Local Similarity 30.4%; Pred. No. 3.3;
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

Qy 6 VAAPAAI-VVGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQYGSANAALALQSDAR 64
Db 99 VAAAPVPAVSTSSAAGVAPNPAGSGNNSPSSSSPTSS-SSSSPSPGSSLAESPEAA 157

Qy 65 KSETTIT---QSGYNGADVGQGDNSTIELTQNG--FRNNATIDQWNAKNSDI 113
Db 158 GVSSTAPLPGAAGPGTGVPAVSGALRELLEACRNGDVSRVKRLVDAANVNAKDM 212

RESULT 13
US-10-199-937-4
; Sequence 4; Application US/10199937
; Publication No. US20030190739A1
; GENERAL INFORMATION:
; APPLICANT: Christenson, Erik
; APPLICANT: DeMaggio, Anthony J.
; APPLICANT: Goldman, Phyllis S.
; APPLICANT: McElligott, David L.
; TITLE OF INVENTION: TANKYRASE2 MATERIALS AND METHODS
; FILE REFERENCE: 27866/36559
; CURRENT APPLICATION NUMBER: US/10/199,937
; CURRENT FILING DATE: 2002-07-22
; PRIOR APPLICATION NUMBER: US/09/606,035
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:48:33 ; Search time 167.9 Seconds
(without alignments)
877.809 Million cell updates/sec

Title: US-09-543-407-14
Perfect score: 775
Sequence: 1 MKLKVAFAAIVVSGSALA.....HASYMVRQVGFNNATANQY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 6019581 seqs, 97603577 residues

Total number of hits satisfying chosen parameters: 6019581

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Parents AA Main:*

1: /cgn2_6/ptodata/2/paa/PCTUS_COMB.pep.*
2: /cgn2_6/ptodata/2/paa/US06_COMB.pep.*
3: /cgn2_6/ptodata/2/paa/US07_COMB.pep.*
4: /cgn2_6/ptodata/2/paa/US08_COMB.pep.*
5: /cgn2_6/ptodata/2/paa/US081_COMB.pep.*
6: /cgn2_6/ptodata/2/paa/US082_COMB.pep.*
7: /cgn2_6/ptodata/2/paa/US083_COMB.pep.*
8: /cgn2_6/ptodata/2/paa/US084_COMB.pep.*
9: /cgn2_6/ptodata/2/paa/US085_COMB.pep.*
10: /cgn2_6/ptodata/2/paa/US086_COMB.pep.*
11: /cgn2_6/ptodata/2/paa/US087_COMB.pep.*
12: /cgn2_6/ptodata/2/paa/US088_COMB.pep.*
13: /cgn2_6/ptodata/2/paa/US089_COMB.pep.*
14: /cgn2_6/ptodata/2/paa/US090_COMB.pep.*
15: /cgn2_6/ptodata/2/paa/US091_COMB.pep.*
16: /cgn2_6/ptodata/2/paa/US092_COMB.pep.*
17: /cgn2_6/ptodata/2/paa/US093_COMB.pep.*
18: /cgn2_6/ptodata/2/paa/US094_COMB.pep.*
19: /cgn2_6/ptodata/2/paa/US095_COMB.pep.*
20: /cgn2_6/ptodata/2/paa/US096_COMB.pep.*
21: /cgn2_6/ptodata/2/paa/US097A_COMB.pep.*
22: /cgn2_6/ptodata/2/paa/US097B_COMB.pep.*
23: /cgn2_6/ptodata/2/paa/US098_COMB.pep.*
24: /cgn2_6/ptodata/2/paa/US099A_COMB.pep.*
25: /cgn2_6/ptodata/2/paa/US099B_COMB.pep.*
26: /cgn2_6/ptodata/2/paa/US100_COMB.pep.*
27: /cgn2_6/ptodata/2/paa/US101_COMB.pep.*
28: /cgn2_6/ptodata/2/paa/US102_COMB.pep.*
29: /cgn2_6/ptodata/2/paa/US103_COMB.pep.*
30: /cgn2_6/ptodata/2/paa/US104_COMB.pep.*
31: /cgn2_6/ptodata/2/paa/US106_COMB.pep.*
32: /cgn2_6/ptodata/2/paa/US107_COMB.pep.*
33: /cgn2_6/ptodata/2/paa/US160_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
------------	-------	-------------	--------	-------	-------------

1	775	100.0	151	19	US-09-543-407-14	Sequence 14, Appl
2	714	92.1	151	19	US-09-543-407-24	Sequence 24, Appl
3	696	89.8	151	19	US-09-543-407-12	Sequence 12, Appl
4	693	89.4	151	19	US-09-543-407-5	Sequence 5, Appl
5	692	89.3	151	6	US-08-233-642A-57	Sequence 57, Appl
6	659	85.0	151	19	US-09-543-407-30	Sequence 30, Appl
7	632	79.0	151	19	US-09-543-407-26	Sequence 26, Appl
8	611	78.8	151	19	US-09-543-407-18	Sequence 18, Appl
9	609	78.6	151	19	US-09-543-407-20	Sequence 20, Appl
10	606	78.2	151	19	US-09-543-407-31	Sequence 31, Appl
11	603	77.8	151	19	US-09-543-407-28	Sequence 28, Appl
12	602	77.7	151	19	US-09-543-407-22	Sequence 22, Appl
13	578	74.6	151	19	US-09-543-407-16	Sequence 16, Appl
14	528	68.1	151	19	US-09-543-407-7	Sequence 7, Appl
15	525	67.7	151	13	US-08-978-878-4	Sequence 4, Appl
16	525	67.7	151	21	US-09-741-873B-4	Sequence 4, Appl
17	523	67.5	151	33	US-60-352-946-2	Sequence 2, Appl
18	523	67.5	151	33	US-60-444-371-2	Sequence 2, Appl
19	514	66.3	120	6	US-08-233-642A-55	Sequence 55, Appl
20	474	61.2	109	19	US-09-543-407-34	Sequence 34, Appl
21	472	60.9	158	16	US-09-252-691-5834	Sequence 5834, Ap
22	472	60.9	158	16	US-09-252-691C-5834	Sequence 5834, Ap
23	472	60.9	158	30	US-10-417-886-5834	Sequence 5834, Ap
24	447	57.7	131	13	US-08-978-878-2	Sequence 2, Appl
25	447	57.7	131	21	US-09-741-873B-2	Sequence 2, Appl
26	347	44.8	68	19	US-09-543-407-37	Sequence 37, Appl
27	343	44.3	109	19	US-09-543-407-32	Sequence 32, Appl
28	243.5	31.4	70	19	US-09-543-407-35	Sequence 35, Appl
29	237	30.6	48	19	US-09-543-407-39	Sequence 39, Appl
30	115.5	14.9	186	16	US-09-253-691-5833	Sequence 5833, Ap
31	115.5	14.9	186	16	US-09-253-691C-5833	Sequence 5833, Ap
32	115.5	14.9	186	30	US-10-417-886-5833	Sequence 5833, Ap
33	104.5	13.5	151	19	US-09-543-407-8	Sequence 8, Appl
34	104	13.4	445	29	US-10-369-493-20638	Sequence 20638, A
35	104	13.4	445	33	US-60-360-039-20638	Sequence 20638, A
36	98.5	12.7	151	19	US-09-543-407-6	Sequence 6, Appl
37	98	12.6	210	28	US-10-219-999-43013	Sequence 43013, A
38	98	12.6	210	30	US-10-425-114-53421	Sequence 53421, A
39	98	12.6	210	30	US-10-425-114A-53421	Sequence 53421, A
40	98	12.6	210	33	US-60-312-544-6765	Sequence 6765, Ap
41	98	12.6	210	33	US-60-324-109-30520	Sequence 30520, A
42	98	12.6	211	30	US-10-424-599-245046	Sequence 245046,
43	95	12.3	186	28	US-10-282-132A-49412	Sequence 49412, A
44	92	11.9	673	20	US-09-678-411-8	Sequence 8, Appl
45	92	11.9	949	20	US-09-678-411-10	Sequence 10, Appl

ALIGNMENTS

RESULT 1
US-09-543-407-14
; Sequence 14, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: PastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b atga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding P73 from GP63 of Leishmania major.

US-09-543-407-14

```
Query Match      100.0%; Score 775; DB 19; Length 151;
Best Local Similarity 100.0%; Pred. No. 6.1e-75;
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQYGSANAALALQ 60
   |||||
Db 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQYGSANAALALQ 60
   |||||

Qy 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITVQYDQ 120
   |||||
Db 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITVQYDQ 120
   |||||

Qy 121 LVTRVVVTHEMAHASVMVRQVGFNNATANQY 151
   |||||
Db 121 LVTRVVVTHEMAHASVMVRQVGFNNATANQY 151
   |||||
```

RESULT 2

```
US-09-543-407-24
; Sequence 24, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-24
```

```
Query Match      92.1%; Score 714; DB 19; Length 151;
Best Local Similarity 91.1%; Pred. No. 2.4e-68;
Matches 144; Conservative 0; Mismatches 0; Indels 14; Gaps 2;

Qy 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQYGSANAALALQ 60
   |||||
Db 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQYGSANAALALQ 60
   |||||

Qy 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITVQYDQ 120
   |||||
Db 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITVQYDQ 113
   |||||

Qy 121 LVTRVVVTHEMAHA-----SYMVRQVGFNNATANQY 151
   |||||
Db 114 LVTRVVVTHEMAHANQTASDSSVMVRQVGFNNATANQY 151
   |||||
```

RESULT 3

```
US-09-543-407-12
; Sequence 12, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
```

```
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-12
```

```
Query Match      89.8%; Score 696; DB 19; Length 151;
Best Local Similarity 87.6%; Pred. No. 2.1e-66;
Matches 141; Conservative 0; Mismatches 0; Indels 20; Gaps 2;

Qy 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQYGSANAALALQ 60
   |||||
Db 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQYGSANAALALQ 60
   |||||

Qy 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITVQYDQ 117
   |||||
Db 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITVQYGG 120
   |||||

Qy 118 -----YDQLVTRVVTHEMAHASVMVRQVGFNNATANQY 151
   |||||
Db 121 NNAALVNYDQLVTRVVTHEMAHA-----NNATANQY 151
   |||||
```

RESULT 4

```
US-09-543-407-5
; Sequence 5, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Salmonella enteritidis
US-09-543-407-5
```

```
Query Match      89.4%; Score 693; DB 19; Length 151;
Best Local Similarity 90.7%; Pred. No. 4.4e-66;
Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

Qy 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQYGSANAALALQ 60
   |||||
Db 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQYGSANAALALQ 60
   |||||

Qy 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITVQYDQ 120
   |||||
Db 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITVQYGG 120
   |||||

Qy 121 LVTRVVVTHEMAHASVMVRQVGFNNATANQY 151
   |||||
Db 121 NNAALVNYQTASDSSVMVRQVGFNNATANQY 151
   |||||
```

RESULT 5

```
US-08-233-642A-57
; Sequence 57, Application US/08233642A
; GENERAL INFORMATION:
; APPLICANT: Kay, William W.
```

```

; APPLICANT: Collinson, S. Karen
; APPLICANT: Clouthier, Sharon C.
; APPLICANT: Doran, James L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
; BASED VACCINES
; NUMBER OF SEQUENCES: 58
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,642A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERY
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 151 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-233-642A-57

Query Match      89.3%; Score 692; DB 6; Length 151;
Best Local Similarity 90.7%; Pred. No. 5,6e-66;
Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYDQ 120
DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYDQ 120
QY 121 LVTRVVTHEMAHASVVRQVGFNNATANQY 151
DB 121 NNPAALVNTASDSSVMVRQVGFNNATANQY 151

RESULT 6
US-09-543-407-30
; Sequence 30, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-30

Query Match      79.0%; Score 612; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 2.5e-57;
Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYDQ 120
DB 61 LVTRVVTHEMAHASVVRQVGFNNATANQY 151
QY 121 LVTRVVTHEMAHASVVRQVGFNNATANQY 151
DB 121 NNAALVNTASDSSVMVRQVGFNNATANQY 151

RESULT 8
US-09-543-407-18
; Sequence 18, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
; BASED VACCINES
; NUMBER OF SEQUENCES: 58
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,642A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERY
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 151 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-233-642A-57

Query Match      85.0%; Score 659; DB 19; Length 151;
Best Local Similarity 81.9%; Pred. No. 2.1e-62;
Matches 136; Conservative 0; Mismatches 0; Indels 30; Gaps 2;

QY 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYDQ 120
DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYDQ 120
QY 121 LVTRVVTHEMAHASVVRQVGFNNATANQY 151
DB 106 LVTRVVTHEMAHAGGNAALVNTASDSSVMVRQVGFNNATANQY 151

RESULT 7
US-09-543-407-26
; Sequence 26, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-26

Query Match      89.3%; Score 692; DB 6; Length 151;
Best Local Similarity 90.7%; Pred. No. 5,6e-66;
Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYDQ 120
DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYDQ 120
QY 121 LVTRVVTHEMAHASVVRQVGFNNATANQY 151
DB 121 NNPAALVNTASDSSVMVRQVGFNNATANQY 151

RESULT 6
US-09-543-407-30
; Sequence 30, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-30

Query Match      85.0%; Score 659; DB 19; Length 151;
Best Local Similarity 81.9%; Pred. No. 2.1e-62;
Matches 136; Conservative 0; Mismatches 0; Indels 30; Gaps 2;

QY 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYDQ 120
DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYDQ 120
QY 121 LVTRVVTHEMAHASVVRQVGFNNATANQY 151
DB 106 LVTRVVTHEMAHAGGNAALVNTASDSSVMVRQVGFNNATANQY 151

RESULT 7
US-09-543-407-26
; Sequence 26, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-26

Query Match      79.0%; Score 612; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 2.5e-57;
Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYDQ 120
DB 61 LVTRVVTHEMAHASVVRQVGFNNATANQY 151
QY 121 LVTRVVTHEMAHASVVRQVGFNNATANQY 151
DB 121 NNAALVNTASDSSVMVRQVGFNNATANQY 151

RESULT 8
US-09-543-407-18
; Sequence 18, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
; BASED VACCINES
; NUMBER OF SEQUENCES: 58
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,642A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERY
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 151 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-233-642A-57
```

```

; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-18

Query Match          78.8%; Score 611; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 3.2e-57;
Matches 123; Conservative 6; Mismatches 22; Indels 0; Gaps 0;

Qy 1 MKLLKVAFAAIVVSSALAGVVPQWGGGNGGNSGGPDSITLSIYQYGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSSALAGVVPQWGGGNGGNSGGPDDYDQLVTRVVTHEMAHALQ 60

Qy 61 SPARKSETTITSGYNGADVQGGADNSTIELTQNGFRNATIDQWNAKNSDITVGYDQ 120
Db 61 SPARKSETTITSGYNGADVQGGADNSTIELTQNGFRNATIDQWNAKNSDITVGYGG 120

Qy 121 LVTRVVTHEMAHASVVMVROVQFGNNATANQY 151
Db 121 NNAALVNQTASDSSVMVROVQFGNNATANQY 151

RESULT 9
US-09-543-407-20
; Sequence 20, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-20

Query Match          78.6%; Score 609; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 5.2e-57;
Matches 123; Conservative 6; Mismatches 22; Indels 0; Gaps 0;

Qy 1 MKLLKVAFAAIVVSSALAGVVPQWGGGNGGNSGGPDSITLSIYQYGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSSALAGVVPQWGGGNGGNSGGPDSITLSIYQYGSANAALALQ 60

Qy 61 SPARKSETTITSGYNGADVQGGADNSTIELTQNGFRNATIDQWNAKNSDITVGYDQ 120
Db 61 SPARKYDQLVTRVVTHEMAHAGGADNSTIELTQNGFRNATIDQWNAKNSDITVGYGG 120

```

```

Qy 121 LVTRVVTHEMAHASVVMVROVQFGNNATANQY 151
Db 121 NNAALVNQTASDSSVMVROVQFGNNATANQY 151

RESULT 10
US-09-543-407-31
; Sequence 31, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Salmonella enteritidis
; US-09-543-407-31

Query Match          78.2%; Score 606; DB 19; Length 151;
Best Local Similarity 89.3%; Pred. No. 9e-57;
Matches 117; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

Qy 21 GVVPQWGGGNGGNSGGPDSITLSIYQYGSANAALALQSDARKSETTITSGYNGAD 80
Db 1 GVVPQWGGGNGGNSGGPDSITLSIYQYGSANAALALQSDARKSETTITSGYNGAD 60

Qy 81 VGGADNSTIELTQNGFRNATIDQWNAKNSDITVGYDQLVTRVVTHEMAHASVVMVROV 140
Db 61 VGGADNSTIELTQNGFRNATIDQWNAKNSDITVGYDQLVTRVVTHEMAHASVVMVROV 120

Qy 141 GFGNNATANQY 151
Db 121 GFGNNATANQY 131

RESULT 11
US-09-543-407-28
; Sequence 28, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-28

Query Match          77.8%; Score 603; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 2.3e-56;
Matches 122; Conservative 4; Mismatches 25; Indels 0; Gaps 0;

Qy 1 MKLLKVAFAAIVVSSALAGVVPQWGGGNGGNSGGPDSITLSIYQYGSANAALALQ 60

```

Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDSITLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGQGDNSTIETQNGFRNNATIDQWNAKNSDITVGYDQ 120
Db 61 SDARKSETTITQSGYNGADVGQGDNSTIETQNGFRNNATIDQWNAKNSDITVGYDQ 120
QY 121 LVTRVVTHEMAHASVVMVROVGFNNATANQY 151
Db 121 NNAALVNTASDSSVMVROVGFNNATANQY 151

RESULT 12

US-09-543-407-22
; Sequence 22, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 22
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-22

Query Match 77.7%; Score 602; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 3e-56;
Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDSITLSIYQYGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDSITLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGQGDNSTIETQNGFRNNATIDQWNAKNSDITVGYDQ 120
Db 61 SDARKSETTITQSGYNGADVGQGDNSTIETQNGFRNNATIDQWNAKNSDITVGYDQ 120
QY 121 LVTRVVTHEMAHASVVMVROVGFNNATANQY 151
Db 121 NNAALVNTASDSSVMVROVGFNNATANQY 151

RESULT 13

US-09-543-407-16
; Sequence 16, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence

; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-16

Query Match 74.6%; Score 578; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 1.2e-53;
Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDSITLSIYQYGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDSITLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGQGDNSTIETQNGFRNNATIDQWNAKNSDITVGYDQ 120
Db 61 SDARKSETTITQSGYNGADVGQGDNSTIETQNGFRNNATIDQWNAKNSDITVGYDQ 120
QY 121 LVTRVVTHEMAHASVVMVROVGFNNATANQY 151
Db 121 NNAALVNTASDSSVMVROVGFNNATANQY 151

RESULT 14

US-09-543-407-7
; Sequence 7, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-543-407-7

Query Match 68.1%; Score 528; DB 19; Length 151;
Best Local Similarity 68.9%; Pred. No. 2.9e-48;
Matches 104; Conservative 20; Mismatches 27; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDSITLSIYQYGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDSITLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGQGDNSTIETQNGFRNNATIDQWNAKNSDITVGYDQ 120
Db 61 TDARNSDLTITQGGNGADVGQGDSDSIDITQRFGNSATLDQWNGKNSMTVYQFGG 120
QY 121 LVTRVVTHEMAHASVVMVROVGFNNATANQY 151
Db 121 GNGAAVDQTASNSVNVTVQVGFNNATANQY 151

RESULT 15

US-08-978-878-4
; Sequence 4, Application US/08978878
; GENERAL INFORMATION:
; APPLICANT: NORMARK, Staffan
; APPLICANT: OLSEN, Arne
; TITLE OF INVENTION: FIBRONECTIN BINDING PROTEIN AS WELL AS ITS PREPARATION
; FILE REFERENCE: 012889-081
; CURRENT APPLICATION NUMBER: US/08/978,878
; CURRENT FILING DATE: 1997-11-26
; EARLIER APPLICATION NUMBER: SE 8801723-1
; EARLIER FILING DATE: 1988-05-06

```
; EARLIER APPLICATION NUMBER: US 07/347,189
; EARLIER FILING DATE: 1989-05-04
; EARLIER APPLICATION NUMBER: US 07/789,437
; EARLIER FILING DATE: 1991-11-06
; EARLIER APPLICATION NUMBER: US 07/970,846
; EARLIER FILING DATE: 1992-11-03
; EARLIER APPLICATION NUMBER: US 08/187,865
; EARLIER FILING DATE: 1994-01-28
; EARLIER APPLICATION NUMBER: US 08/318,519
; EARLIER FILING DATE: 1994-10-05
; EARLIER APPLICATION NUMBER: US 08/495,959
; EARLIER FILING DATE: 1995-06-28
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-08-978-878-4

Query Match      67.7%; Score 525; DB 13; Length 151;
Best Local Similarity 68.2%; Pred. No. 6.1e-48;
Matches 103; Conservative 21; Mismatches 27; Indels 0; Gaps 0;

QY      1 MKLLKVAAPAAIVVSGSALAGVVPQVGGGNGHGGNNGSGDPSTLSIYQGSANAALALQ 60
DB      1 MKLLKVAATAAIVFSGSAGVVPQVGGGNGHGGNNGSGPNSELNIYQGGNSALALQ 60

QY      61 SDARKSETTITQSGYNGADVGCGADNSTIELTQNGFRNNATIDOWNAKNSDITVGYDQ 120
DB      61 TDARNSDLTITQHGCGNGADVGCGSDSDSIDLTQRFNGNSATLDQNGKNSMTVKQFEG 120

QY      121 LVTRVVTHEMAHASVMVROVFGNNATANQY 151
DB      121 GNGAAVDQTASNSSVNVTVQVFGNNATAHQY 151

Search completed: August 2, 2004, 15:26:42
Job time : 167.9 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:49:38 ; Search time 17.8 Seconds
(without alignments)
888.146 Million cell updates/sec

Title: US-09-543-407-14
Perfect score: 775
Sequence: 1 MKLLKVAFAAIVVSGSALA.....HASVMRVQVFGNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 601315 seqs, 104695340 residues

Total number of hits satisfying chosen parameters: 601315

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents AA New:*

1: /cgn2_6/ptodata/2/paa/PCT NEW COMB.pdb:*
2: /cgn2_6/ptodata/2/paa/US06 NEW COMB.pdb:*
3: /cgn2_6/ptodata/2/paa/US07 NEW COMB.pdb:*
4: /cgn2_6/ptodata/2/paa/US08 NEW COMB.pdb:*
5: /cgn2_6/ptodata/2/paa/US09 NEW COMB.pdb:*
6: /cgn2_6/ptodata/2/paa/US10 NEW COMB.pdb:*
7: /cgn2_6/ptodata/2/paa/US60 NEW COMB.pdb:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	525	67.7	151	US-09-741-873C-4	Sequence 4, Appli
2	447	57.7	131	US-09-741-873C-2	Sequence 2, Appli
3	92	11.9	299	US-10-170-205E-35751	Sequence 35751, A
4	92	11.9	1203	US-10-170-205E-741	Sequence 741, Appl
5	92	11.9	1327	PCT-US04-02338-49	Sequence 49, Appl
6	90.5	11.7	258	US-10-425-115-300390	Sequence 300390,
7	89.5	11.5	256	US-10-425-115-301334	Sequence 301334,
8	89	11.5	295	US-10-425-115-312468	Sequence 312468,
9	88	11.4	520	US-10-479-638-21	Sequence 21, Appl
10	86.5	11.2	527	US-10-468-356-348	Sequence 348, App
11	86	11.1	179	US-10-425-115-346132	Sequence 346132,
12	85.5	11.0	1048	PCT-US04-05654-2086	Sequence 2086, Ap
13	82.5	10.6	535	PCT-US04-21492-88	Sequence 88, Appl
14	82	10.6	758	US-60-565-632-9194	Sequence 9194, Ap
15	82	10.6	758	US-60-565-632-9194	Sequence 9194, Ap
16	82	10.6	1095	PCT-US03-24982A-273	Sequence 273, App
17	82	10.6	2228	US-10-745-237-96	Sequence 96, Appl
18	81.5	10.5	376	US-10-491-733-2	Sequence 2, Appli
19	81.5	10.5	508	US-10-425-115-285216	Sequence 285216,
20	81.5	10.5	532	US-10-425-115-285214	Sequence 285214,
21	81	10.5	131	US-60-565-632-11109	Sequence 11109, A
22	81	10.5	131	US-60-579-062-11109	Sequence 11109, A
23	81	10.5	201	US-10-425-115-309662	Sequence 309662,
24	81	10.5	376	US-10-425-115-342526	Sequence 342526,
25	80.5	10.4	246	US-10-854-439-511	Sequence 511, App
26	80.5	10.4	388	US-09-248-796A-17306	Sequence 17306, A

US-09-741-873C-4
; Sequence 4; Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873C-4

ALIGNMENTS

RESULT 1

Query Match 57.7%; Score 525; DB 5; Length 151;
Best Local Similarity 68.2%; Pred. No. 2.3e-38;
Matches 103; Conservative 21; Mismatches 27; Indels 0; Gaps 0;
Qy 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGGNSGSPDSTLSIYQVGSANAALQ 60
Db 1 MKLLKVAFAAIVFSSAVAGVVPQYGGGNGGNGGNSGPNSELNIYQVGGNSALQ 60
Qy 61 SDARKSETTITQSGYNGADVGQGANSTIELTQNGFRNATIDQNAKNSDITVQYDQ 120
Db 61 TDARNSDLATITQGGGNGADVGQGDSSIDLTQRFNGSATLDQNGKNSMTVKQFG 120
Qy 121 LVTRVTVTHEMAHSAVMVRQVFGNATANQY 151
Db 121 NGGAANDQTASNSVNTQVFGNATANQY 151

```
RESULT 2
US-09-741-873C-2
; Sequence 2, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Olsark, Staffan
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873C-2

Query Match 57.7%; Score 447; DB 5; Length 131;
Best Local Similarity 64.9%; Pred. No. 1.2e-31;
Matches 85; Conservative 20; Mismatches 26; Indels 0; Gaps 0;

QY 21 GVVPPQGGGNGHNGGNSGPDSTLSIYQGSANAALALQSDAR 80
DB 1 GVVPPQGGGNGHNGGNSGPDSTLSIYQGSANAALALQSDAR 60
QY 81 VGGGADNSTIETONGFRNATIDOWNAKNSDITGVQYDQLVTRVVTHEMAHASVWVQV 140
DB 61 VGGGSDSSIDLTQRFNGSAILDQNGKNSMTVKQFGGNGAAVDQTASNVVTVQV 120
QY 141 GFGNNTATQY 151
DB 121 GFGNNTATQY 131

RESULT 3
US-10-170-205E-35751
; Sequence 35751, Application US/10170205E
; GENERAL INFORMATION:
; APPLICANT: ADAMS, Mark
; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN
; FILE REFERENCE: CL001381
; CURRENT APPLICATION NUMBER: US/10/170,205E
; CURRENT FILING DATE: 2002-06-13
; NUMBER OF SEQ ID NOS: 40312
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 35751
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-170-205E-35751

Query Match 11.9%; Score 92; DB 6; Length 299;
Best Local Similarity 30.4%; Pred. No. 2.2;
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

QY 6 VAAFAAI-VVGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQGSANAALALQSDAR 64
DB 99 VAAAPVPAVSTSSAAGVAPNPAGSGNNSPSSSPTSS-SSSSPSPGSLAESPEAA 157
```

```
DB 99 VAAAPVPAVSTSSAAGVAPNPAGSGNNSPSSSPTSS-SSSSPSPGSLAESPEAA 157
QY 65 KSETTIT----QSGYNGADVGQADNSTIETLQNG--FRNNATIDOWNAKNSDI 113
DB 158 GVSSTAPLPGGAAGPGTGVPAVSGALRELLEACRNGDVSXKRLVDAANVNAKDM 212

RESULT 4
US-10-170-205E-741
; Sequence 741, Application US/10170205E
; GENERAL INFORMATION:
; APPLICANT: ADAMS, Mark
; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN
; FILE REFERENCE: CL001381
; CURRENT APPLICATION NUMBER: US/10/170,205E
; CURRENT FILING DATE: 2002-06-13
; NUMBER OF SEQ ID NOS: 40312
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 741
; LENGTH: 1203
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-170-205E-741

Query Match 11.9%; Score 92; DB 6; Length 1203;
Best Local Similarity 30.4%; Pred. No. 11;
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

QY 6 VAAFAAI-VVGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQGSANAALALQSDAR 64
DB 99 VAAAPVPAVSTSSAAGVAPNPAGSGNNSPSSSPTSS-SSSSPSPGSLAESPEAA 157
QY 65 KSETTIT----QSGYNGADVGQADNSTIETLQNG--FRNNATIDOWNAKNSDI 113
DB 158 GVSSTAPLPGGAAGPGTGVPAVSGALRELLEACRNGDVSXKRLVDAANVNAKDM 212

RESULT 5
PCT-US04-02338-49
; Sequence 49, Application PC/TUS0402338
; GENERAL INFORMATION:
; APPLICANT: EXELIXIS, INC.
; TITLE OF INVENTION: USE
; FILE REFERENCE: EX04-003C-PC
; CURRENT APPLICATION NUMBER: PCT/US04/02338
; CURRENT FILING DATE: 2004-01-28
; PRIOR APPLICATION NUMBER: US60/443,484
; PRIOR FILING DATE: 2003-01-29
; PRIOR APPLICATION NUMBER: US60/447,358
; PRIOR FILING DATE: 2003-02-11
; PRIOR APPLICATION NUMBER: US60/461,789
; PRIOR FILING DATE: 2003-04-10
; PRIOR APPLICATION NUMBER: US60/470,684
; PRIOR FILING DATE: 2003-05-14
; PRIOR APPLICATION NUMBER: US60/479,650
; PRIOR FILING DATE: 2003-06-19
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 49
; LENGTH: 1327
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US04-02338-49

Query Match 11.9%; Score 92; DB 1; Length 1327;
Best Local Similarity 30.4%; Pred. No. 12;
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

QY 6 VAAFAAI-VVGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQGSANAALALQSDAR 64
DB 99 VAAAPVPAVSTSSAAGVAPNPAGSGNNSPSSSPTSS-SSSSPSPGSLAESPEAA 157
```

QY 65 KSETTIT---QSGYNGADYQGADNSTIELTQNG--FRNATIDQWNAKNSDI 113
Db 158 GVSSITAPLPGAGGTGVPVSGALRELLEACRNGDVSRVKRLVDAANVNAKDM 212

RESULT 6
US-10-425-115-300390
; Sequence 300390, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 300390
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Zea mays
; NAME/KEY: unsure
; LOCATION: (1):(258)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_37025C.1.pap
US-10-425-115-300390

Query Match 11.7%; Score 90.5; DB 6; Length 258;
Best Local Similarity 27.9%; Pred. No. 2.5;
Matches 29; Conservative 15; Mismatches 39; Indels 21; Gaps 4;

QY 17 SALAGVWPQWGGGNGHNGSSGPDSTLSIYQGSANAALQSDARKSETTITQSG-- 74
Db 78 SSIAG-----GGGGQGGGGTNGSGSGGGSGYSGSSSTAASGPGSNGVADAEKGGAG 132

QY 75 -----YNGA--DVQGANSTIEL--TONGFRNATID 104
Db 133 GGMGGGANGAYGSGAGGKGEGVSGVALAPSSNGYNGGAAD 176

RESULT 7
US-10-425-115-301334
; Sequence 301334, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 301334
; LENGTH: 256
; TYPE: PRT
; ORGANISM: Zea mays
; NAME/KEY: unsure
; LOCATION: (1):(256)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_37894C.1.pap
US-10-425-115-301334

Query Match 11.5%; Score 89.5; DB 6; Length 256;
Best Local Similarity 27.9%; Pred. No. 3;
Matches 29; Conservative 14; Mismatches 40; Indels 21; Gaps 4;

QY 17 SALAGVWPQWGGGNGHNGSSGPDSTLSIYQGSANAALQSDA---RKSETTITQ 72
Db 133 GGMGGGANGAYGSGAGGKGEGVSGVALAPSSNGYNGGAAD 176

Db 79 SSVAG-----GGGGGGGGGGTNGSGSGGGSGYSGTSTTAASGPPSSGNYANAEKGAG 133
QY 73 SGYNGAD-----VGQGANSTIEL--TONGFRNATID 104
Db 134 GGMGGGADGAYGSGAGGKGEGVSGVALAPSSDGYNGGAAD 177

RESULT 8
US-10-425-115-312468
; Sequence 312468, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 312468
; LENGTH: 295
; TYPE: PRT
; ORGANISM: Zea mays
; NAME/KEY: unsure
; LOCATION: (1):(295)
; OTHER INFORMATION: Clone ID: MRT4577_48027C.1.pap
US-10-425-115-312468

Query Match 11.5%; Score 89; DB 6; Length 295;
Best Local Similarity 35.1%; Pred. No. 3.9;
Matches 33; Conservative 10; Mismatches 39; Indels 12; Gaps 4;

QY 26 WGGGNGHNGGSSGPDSTLSIYQGSANAALQSDARKSETTITQSGYNGADYQGA 85
Db 138 YGGGYSYSGGSSG-GYAANGYGVSGSGNYNAGSGGYSGS----DGYNGAASGGYA 192

QY 86 DNSTIELTQNGFRNATIDQWNAKNSDITVGQYD 119
Db 193 NNLSGYNNGRYN--TIG-----SSDGNTEGYN 219

RESULT 9
US-10-479-638-21
; Sequence 21, Application US/10479638
; GENERAL INFORMATION:
; APPLICANT: Don A. Reth
; APPLICANT: Randolph V. Lewis
; TITLE OF INVENTION: Expression of Spider Silk Proteins in Higher Plants
; FILE REFERENCE: WYO.02-0004US
; CURRENT APPLICATION NUMBER: US/10/479,638
; CURRENT FILING DATE: 2003-12-03
; PRIOR APPLICATION NUMBER: PCT/US02/18256
; PRIOR FILING DATE: 2002-06-06
; PRIOR APPLICATION NUMBER: 60/296,184
; PRIOR FILING DATE: 2001-06-06
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 21
; LENGTH: 520
; TYPE: PRT
; ORGANISM: Argiope trifasciata
US-10-479-638-21

Query Match 11.4%; Score 88; DB 6; Length 520;
Best Local Similarity 21.2%; Pred. No. 9.2;
Matches 35; Conservative 30; Mismatches 66; Indels 38; Gaps 6;

QY 15 SGSALAGVWPQWGGGNGHNGSSGPDSTLSIYQGSANAALQSDARKSETTITQ 73
Db 347 AGAAAAAASAGAGGAGGCGGCVAGGSS-----ISYCATSSSATSSSTASSRSRGTG 402

QY 74 GYNGADVGGADNSTIELTONG-FRNATIDOWNAKNSDI----- 113
DB 403 GYGAGAAAGAGAAAGAGSISRLSSAAVAVRVSNICAVASGASALPGVISNIF 462
QY 114 -----TVGYDQIVTR-----VVTHEMAHASV-MVRQVGFNNATNQ 150
DB 463 SGVSSAGSYEEAVIQSLLEVLVLLHILSNSSIGYVGAEGUGDSIAVVQ 512

RESULT 10
US-10-468-356-348
; Sequence 348, Application US/10468356
; GENERAL INFORMATION:
; APPLICANT: COLE, STEWART
; TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENEOMICS AS A TOOL FOR
; IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR
; TREATMENT OF MYCOBACTERIOSES
; FILE REFERENCE: 05394.0019
; CURRENT APPLICATION NUMBER: US/10/468,356
; PRIOR FILING DATE: 2003-08-19
; PRIOR APPLICATION NUMBER: 10/080,170
; PRIOR FILING DATE: 2002-02-22
; PRIOR APPLICATION NUMBER: 60/270,123
; PRIOR FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 655
; SOFTWARE: Patent In Ver. 3.2
; SEQ ID NO 348
; LENGTH: 527
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-10-468-356-348

Query Match 11.2%; Score 86.5; DB 6; Length 527;
Best Local Similarity 22.9%; Pred. No. 13;
Matches 35; Conservative 13; Mismatches 56; Indels 49; Gaps 5;

QY 21 GVVPWGG---CGNHGGGNSGPD-----STLSIYOGSANAALALQSPARKSE 67
DB 387 GYAPQGGGYAEPAGRDYDYGSGADPDYQGPAGPGYSGYSGAGTSVTLQLDDGSGR 446
QY 68 TTITQSG---YNGADV-----GGADNSTIELTQNGFRNNATIDQ 105
DB 447 TVQLREGSNIIGQDQAQPLDPTGVSRRLHBIWDGQVALLADLNSGTGTVNNAVQE 506
QY 106 WNAKNSDIYGVQDQLVTRVTHEMAHASVMVR 138
DB 507 WQADGDVT-----RLGHSIIVR 525

RESULT 11
US-10-425-115-346132
; Sequence 346132, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; Fungal Pathogenesis
; FILE REFERENCE: 38-21(S3222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; PRIOR FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 346132
; LENGTH: 179
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_78839C.1.pap
US-10-425-115-346132

Query Match 11.1%; Score 86; DB 6; Length 179;
Best Local Similarity 29.9%; Pred. No. 4.1;

Matches 32; Conservative 11; Mismatches 40; Indels 24; Gaps 5;
QY 19 LAGVVPWGGG-----NHNGGNSGPDSTLSIYOGSANAALALQSDARKSETTI 70
DB 64 LDGLLGLGGGLDGLGLTGGGKKNKAQADSGNAEGGNA-----QEDSGNAQ--- 116
QY 71 TQSGYNGADVGGADNSTIELTQNGFRNNATIDOWNAKNSDIYVQ 117
DB 117 EESGNQACAGAGA-----ENGAANNGT-EAAGAENAAAGNQ 154

RESULT 12
PCT-US04-05654-2086
; Sequence 2086, Application PC/TUS0405654
; GENERAL INFORMATION:
; APPLICANT: Sherman, Bradley K
; APPLICANT: Riechmann, Jose Luis
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Heard, Jacqueline E
; APPLICANT: Haake, Volker
; APPLICANT: Creelman, Robert A
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Adam, Luc J
; APPLICANT: Reuber, T. Lynne
; APPLICANT: Keddie, James
; APPLICANT: Dubell III, Arnold N
; APPLICANT: Pineda, Omaira
; APPLICANT: Repetti, Peter
; APPLICANT: Century, Karen
; APPLICANT: Gutterson, Neal
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Broun, Pierre E
; APPLICANT: Kumimoto, Roderick W
; APPLICANT: Pilgrim, Marsha L
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES IN PLANTS
; FILE REFERENCE: MBI-0047 PCT
; CURRENT APPLICATION NUMBER: PCT/US04/05654
; CURRENT FILING DATE: 2004-03-03
; PRIOR APPLICATION NUMBER: 10/374,780
; PRIOR FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: 10/675,852
; PRIOR FILING DATE: 2003-09-30
; NUMBER OF SEQ ID NOS: 2950
; SOFTWARE: Patent In version 3.2
; SEQ ID NO 2086
; LENGTH: 1048
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: G1152 Paralogous to G1146
PCT-US04-05654-2086

Query Match 11.0%; Score 85.5; DB 1; Length 1048;
Best Local Similarity 25.7%; Pred. No. 34;
Matches 36; Conservative 17; Mismatches 60; Indels 27; Gaps 3;

QY 27 GGGGNGGNSGSPDSTLSI-----YQGSANAALALQSDARKSETTITQSGYG 76
DB 95 GGYGGGGGSGSPGPPQPSVPELHOATSTPTQAVSSQPTLSEVSPQVPEPTVLAQQF- 153
QY 77 NGADVCGGADNSTIE-----LTQNGFRNNATIDOWNAKNSDIYVQYDQ 120
DB 154 EQLSVEQGPSQAIOPIPSKAFKPFMPRPGKQSGKRCIVKANHFPAELPKDLHHYDV 213
QY 121 LVTRVVTHEMAHASVMVRV 140
DB 214 TITPEVTSRGVNRVANKQLV 233

RESULT 13
PCT-US04-21492-88
; Sequence 88, Application PC/TUS0421492
; GENERAL INFORMATION:

QY 121 LVTRVTHEMAHASVMVRQVGFENNATANQY 151

Db 121 NNAALVNQASDSSVMVRQVGFNNATANQY 151

RESULT 2

AI0635
C:Species: Salmonella enterica subsp. enterica serovar Typhimurium
A:Note: this species has also been called Salmonella typhimurium
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 19-Nov-2002
C:Accession: AI0635
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Garra, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar AB0502; PMID:21534947; PMID:11677608
A:Reference number: AB0502; PMID:21534947; PMID:11677608
A:Accession: AI0635
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-151 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD08268.1; PID:gl6502315; GSPDB:GN00176
C:Genetics:
A:Gene: STY1181

Query Match 89.4%; Score 693; DB 2; Length 151;
Best Local Similarity 90.7%; Pred. No. 5e-52;
Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGPDSSTLSIYYGSAANAALALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGPDSSTLSIYYGSAANAALALQ 60
QY 61 SPARKSETTITQSGYNGADVGADNNTIELTQNGFRNATIDOWNAKNSDITVGOYDQ 120
DB 61 SPARKSETTITQSGYNGADVGADNNTIELTQNGFRNATIDOWNAKNSDITVGOYDQ 120
QY 121 LVTRVVTHMAHASVMVRQVGFNNATANQY 151
DB 121 NNAALVNQASDSSVMVRQVGFNNATANQY 151

RESULT 3

S70788
C:Species: Escherichia coli (strain K-12)
C:Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 01-Mar-2002
C:Accession: S70788; G64846; S31202; S34560; S34559
R:Hammar, M.; Arngvist, A.; Bian, Z.; Olsen, A.; Normark, S.
Mol. Microbiol. 18, 661-670, 1995
A:Title: Expression of two csg operons is required for production of fibronectin- and C:Reference number: S70788; PMID:96414468; PMID:8817489
A:Accession: S70788
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-151 <HAM>
A:Cross-references: EMBL:X90754; NID:91147558; PIDN:CAAG2282.1; PID:gl147564
A:Experimental source: strain K12, substrain W3110
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, D.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; PMID:97426617; PMID:9278503
A:Accession: G64846
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-151 <BLAT>
A:Cross-references: GB:AE000205; GB:U00096; NID:gl787265; PIDN:AAC74126.1; PID:gl787279;
A:Experimental source: strain K-12, substrain MG1655
R:Olsen, A.; Arngvist, A.; Hammar, M.; Sukupolvi, S.; Normark, S.
Mol. Microbiol. 7, 523-536, 1993

A:Title: The RpoS sigma factor relieves H-NS-mediated transcriptional repression of csgA
A:Reference number: S31202; PMID:93211294; PMID:8459772
A:Accession: S31202
A:Molecule type: DNA
A:Residues: 1-6,'V',8-151 <OLS1>
A:Cross-references: EMBL:L04979
A:Accession: S34560
A:Molecule type: protein
A:Residues: 21-42;44-50 <OLS2>
R:Olsen, A.N.; Arngvist, A.N.
submitted to the EMBL Data Library, October 1992
A:Reference number: S34559
A:Accession: S34559
A:Molecule type: DNA
A:Residues: 1-133,'RQDSGMLW' <OLS3>
A:Cross-references: EMBL:L04979; NID:G290424; PIDN:AAA23616.1; PID:G290425
A:Experimental source: strain K-12, substrain W3110
C:Genetics:
A:Gene: csgA
A:Map position: 23.15
C:Function:
A:Description: major component of wild-type curli; interaction between CsgA and CsgB tr:
A:Note: curli are thin, coiled fibers expressed on the surface of Escherichia coli that
and H-20 domain; in the absence of CsgA, CsgB can self-assemble into polymers
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-151/Product: curlin #status experimental <MAT>

Query Match 68.1%; Score 528; DB 2; Length 151;
Best Local Similarity 68.9%; Pred. No. 5.4e-38;
Matches 104; Conservative 20; Mismatches 27; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGPDSSTLSIYYGSAANAALALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGPDSSTLSIYYGSAANAALALQ 60
QY 61 SPARKSETTITQSGYNGADVGADNNTIELTQNGFRNATIDOWNAKNSDITVGOYDQ 120
DB 61 SPARKSETTITQSGYNGADVGADNNTIELTQNGFRNATIDOWNAKNSDITVGOYDQ 120
QY 121 LVTRVVTHMAHASVMVRQVGFNNATANQY 151
DB 121 NNAALVNQASDSSVMVRQVGFNNATANQY 151

RESULT 4

D90806
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: D90806
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen:
A:Reference number: A99629; PMID:21156231; PMID:11258796
A:Accession: D90806
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-152 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA34843.1; PID:gl3360880; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RIMD 050952
C:Genetics:
A:Gene: ECs1420

Query Match 65.4%; Score 506.5; DB 2; Length 152;
Best Local Similarity 67.1%; Pred. No. 3.6e-36;
Matches 102; Conservative 21; Mismatches 28; Indels 1; Gaps 1;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQW-GGGGNGGNGSGPDSSTLSIYYGSAANAALAL 59
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGPDSSTLSIYYGSAANAALAL 60
QY 60 QSDARKSETTITQSGYNGADVGADNNTIELTQNGFRNATIDOWNAKNSDITVGOYD 119

Db 61 QADARNSLTTTQHGGNGGADVGGGSDSSIDLTORFGNSATLDQWNGKDSHMTVXQFG 120

QY 120 QLVTRVTHMAHASVVMVRQVGFNGNATAHQY 151

Db 121 GGNGAAMDQASNSTVNTVQVFGNATAHQY 152

RESULT 5

H85665

hypothetical protein csgA [imported] - Escherichia coli (strain O157:H7, substrain EDL93

C:Species: Escherichia coli

C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001

C:Accession: H85665

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew

iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,

Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: H85665

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-152 <STO>

A:Cross-references: GB:AB005174; NID:g12514574; PIDN:AA055788.1; GSPDB:GN00145; UWGP:Z16

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: csgA

Query Match 65.4%; Score 506.5; DB 2; Length 152;

Best Local Similarity 67.1%; Pred. No. 3.6e-36;

Matches 102; Conservative 21; Mismatches 28; Indels 1; Gaps 1;

QY 1 MKLLKVAFAIVVSGSALAGVVPQW-CGGNHNHGGNSGPDSTLSIYQYGSANAALAL 59

Db 1 MKLLKVAFAIVVSGSALAGVVPQYGGGNGHGGNSGPNSELNIYQYGGNSALAL 60

QY 60 QSDAKSTTTTQSGYNGADVGGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGYD 119

Db 61 QADARNSLTTTQHGGNGGADVGGGSDSSIDLTORFGNSATLDQWNGKDSHMTVXQFG 120

QY 120 QLVTRVTHMAHASVVMVRQVGFNGNATAHQY 151

Db 121 GGNGAAMDQASNSTVNTVQVFGNATAHQY 152

RESULT 6

S70787

curlin nucleator protein csgB precursor - Escherichia coli (strain K-12)

N:Alternate names: csgB protein; curlin nucleation component; minor curlin protein

C:Species: Escherichia coli

C:Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 01-Mar-2002

C:Accession: S70787; F64846

R:Hammar, M.; Arngvist, A.; Bian, Z.; Olsen, A.; Normark, S.

Mol. Microbiol. 18, 661-670, 1995

A:Title: Expression of two csg operons is required for production of fibronectin- and Co

A:Reference number: S70783; MUID:96414468; PMID:8817489

A:Accession: S70787

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-151 <HMS>

A:Cross-references: EMBL:X90754; NID:g1147558; PIDN:CAA62281.1; PID:g1147563

A:Experimental source: strain K12, substrain W3110

A>Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co

.A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A67420; MUID:197426617; PMID:9278503

A:Accession: F64846

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-151 <BLAT>

A:Cross-references: GB:AE000205; GB:U00096; NID:g1787265; PIDN:AA074125.1; PID:g1787278;

A:Experimental source: strain K-12, substrain MG1655

C:Genetics:

A:Gene: csgB

A:Map position: 23.15

C:Function:

A:Description: minor component of wild-type curli; interaction between CsgA and CsgB tr

A>Note: curli are thin, coiled fibers expressed on the surface of Escherichia coli that

and H-kininogen; in the absence of CsgA, CsgB can self-assemble into polymers

F:1-21/Domain: signal sequence #status predicted <SIG>

F:22-151/Product: minor curlin chain #status predicted <MAT>

Query Match 13.5%; Score 104.5; DB 2; Length 151;

Best Local Similarity 29.2%; Pred. No. 0.054;

Matches 35; Conservative 19; Mismatches 59; Indels 7; Gaps 4;

QY 9 FFAIVVSGSAL--AGVVPQWGGGNGHGGNSGPDSTLSIY-QYGSANAALALQSDARK 65

Db 33 FAVNELSKSFNQAAII--GQAGTNNSAQLRGGSKLLAVVAQEGSSNRA-KIDQTGDY 88

QY 66 SETTITQSGYNGADVGGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGYDQVLVTRV 125

Db 89 NLAYIDQASANDASISQAGYGNNTAMIIQKSGNKANITQYGTQKTAIVVQSQSQAIRV 148

RESULT 7

C90806

minor curlin subunit precursor CsgB [imported] - Escherichia coli (strain O157:H7, subs

C:Species: Escherichia coli

C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001

C:Accession: C90806

R:Hayaishi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G

gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen

A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: C90806

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-151 <HAY>

A:Cross-references: GB:BA000007; PIDN:BA034842.1; PID:g133360879; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RMD 0509952

C:Genetics:

A:Gene: Ecs1419

Query Match 13.5%; Score 104.5; DB 2; Length 151;

Best Local Similarity 29.2%; Pred. No. 0.054;

Matches 35; Conservative 19; Mismatches 59; Indels 7; Gaps 4;

QY 9 FFAIVVSGSAL--AGVVPQWGGGNGHGGNSGPDSTLSIY-QYGSANAALALQSDARK 65

Db 33 FAVNELSKSFNQAAII--GQAGTNNSAQLRGGSKLLAVVAQEGSSNRA-KIDQTGDY 88

QY 66 SETTITQSGYNGADVGGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGYDQVLVTRV 125

Db 89 NLAYIDQASANDASISQAGYGNNTAMIIQKSGNKANITQYGTQKTAIVVQSQSQAIRV 148

RESULT 8

G85665

curlin minor chain precursor, CsgA homolog [imported] - Escherichia coli (strain O157:H

C:Species: Escherichia coli

C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001

C:Accession: G85665

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhe

iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca

Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: G85665

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-151 <STO>

A:Cross-references: GB:AE005174; NID:g12514573; PIDN:AA055787.1; GSPDB:GN00145; UWGP:Z1

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: cs9B

Query Match 13.5%; Score 104.5; DB 2; Length 151;
Best Local Similarity 29.2%; Pred. No. 0.054;
Matches 35; Conservative 19; Mismatches 59; Indels 7; Gaps 4;

QY 9 FRAIVVSGAL--AGVVPWGGGNGHNGSSGPDSTLSIY-OYGSANAALALQSDARK 65
DB 33 FAVNELSKSFNOAII---GAGTNNASQALQGGSKLLAVVAQEGSSNRA-KIDQTGDY 88
QY 66 SETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYDQVLT 125
DB 89 NLAYIDQASANDASISQAGYNTAMIIQKSGENKANIYGTQKTAIVVQSQMAIRV 148

RESULT 9

JC6040

fimbria protein agfB precursor - Salmonella enteritidis

C:Species: Salmonella enteritidis

C>Date: 31-Dec-1996 #sequence_revision 31-Dec-1996 #text_change 08-Oct-1999

C:Accession: J06040

R:Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Baner, P.A.; Kay, W.W.

J. Bacteriol. 178, 662-667, 1996

A:Title: Salmonella enteritidis agfBAC operon encoding thin, aggregative fimbriae.

A:Reference number: JC6039; MUID:96146512; PMID:8550497

A:Accession: J06040

A:Molecule type: DNA

A:Residues: 1-151 <COL>

A:Cross-references: GB:U43280; NID:gl184712; PIDN:ARC43598.1; PID:gl184713

A:Experimental source: strain 276755-3b

C:Genetics:

A:Gene: agfB

C:Function:

A:Description: minor component of thin aggregative fimbriae

A:Note: fimbriae bind to fibronectin, plasminogen, tissue plasminogen activator

C:Keywords: fimbria

F:1-21/Domain: signal sequence #status predicted <SIG>

F:22-151/Product: fimbria protein agfB #status predicted <WAT>

Query Match 12.7%; Score 98.5; DB 2; Length 151;
Best Local Similarity 26.2%; Pred. No. 0.17;
Matches 34; Conservative 17; Mismatches 42; Indels 37; Gaps 4;

QY 10 ARAIVVSGALAGVVPWGGGNGHNGSSGPDSTLSIYQYGSANAALALQSDARKSETT 69

DB 58 ARVRQEGSKLLSVISO--EGGNRAKVDQAGNYFNAYIEQTGNAN-----DAS 103

QY 70 ITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYDQVLT 129

DB 104 ISQSAVGNBA-----AIIQKSGENKANIYGTQK-----TAVVQK 140

QY 130 MAHASVMVRQ 139

DB 141 QSHMAIRVQ 150

RESULT 10

D70604

probable PPE protein - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000

C:Accession: D70604

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: D70604

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-552 <COL>

A:Cross-references: GB:Z92774; GB:AL123456; NID:g3261729; PIDN:CAB07133.1; PID:gl877289

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: PPE

C:Superfamily: Phaseolus glycine-rich cell wall protein 1.8

Query Match 12.5%; Score 96.5; DB 2; Length 552;
Best Local Similarity 28.5%; Pred. No. 1.1;
Matches 41; Conservative 10; Mismatches 62; Indels 31; Gaps 6;

QY 15 SGSALAGV-----VPQWGGGNGHNGSSGPD---STLSIYQYGSANAALALQSDA 63

DB 353 SSGSNIGVFNTGANTLVP---GDLNLLGVGNSGANIFGNAGVLTNTFGNASILNTGLG 409

QY 64 RKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYDQVLT 123

DB 410 NAGE---LNTGFGNAGFVNTGFDNSGNVNTGNSGNINTGWNAGNVNTGFG----- 459

QY 124 RVVTHEMAHASVMVRQVGFQGNAT 147

DB 460 -LIITDSGLTNS-----GFGNTCT 476

RESULT 11

T51024

N:Alternate names: protein B7F21.50

C:Species: Neurospora crassa

C>Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000

C:Accession: T51024

R:Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura

submitted to the Protein Sequence Database, July 2000

A:Reference number: Z25286

A:Accession: T51024

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-770 <SCH>

A:Cross-references: EMBL:AJ389901; GSPDB:GN00116; NCSP:B7F21.50

A:Experimental source: BAC clone B7F21; strain OR74A

C:Genetics:

A:Gene: NCSP:B7F21.50

A:Map position: 6

A:Introns: 117/1

Query Match 12.4%; Score 96; DB 2; Length 770;
Best Local Similarity 22.8%; Pred. No. 1.8;
Matches 31; Conservative 19; Mismatches 54; Indels 32; Gaps 3;

QY 28 GCGNHNGGSSGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGADVGQADN 87

DB 389 GAGNNNGSELSPDPNRYFD---GASPPIPSSSHSTTSTTYAGGHNGHGMRESVDS 444

QY 88 STIEL-----TQNGFRNNATIDQWNAKNSDITVGYDQVLT----- 122

DB 445 QSTGLGVHYSSTTTRNQHHWRNQSNASELSADGSEITHGVASPLVGGSSHARGASGGTS 504

QY 123 ---TRVVTHEMAHASV 135

DB 505 YRYTHTHSHSHSGL 520

RESULT 12

S42136

CnJB protein - Tetrahymena thermophila

C:Species: Tetrahymena thermophila

C>Date: 19-Mar-1997 #sequence_revision 18-Jul-1997 #text_change 07-Dec-1999

C:Accession: S42136; S03650

R:Taylor, P.M.; Martindale, D.W.

submitted to the EMBL Data Library, October 1992

A:Reference number: S42136

A:Accession: S42136

A:Molecule type: DNA
A:Residues: 1-1748 <RAY>
A:Cross-references: EMBL:L03710; NID:g161751; PID:g161752
R:Taylor, F.M.; Martindale, D.W.
Nucleic Acids Res. 21, 4610-4614, 1993
A:Title: Retroviral-type zinc fingers and glycine-rich repeats in a protein encoded by c
A:Reference number: S42135; MUID:94051569; PMID:8233798
A:Accession: S42135
A:Molecule type: DNA
A:Residues: 1164-1174;1179-1198;1233-1252;1285-1293;1297-1309;1316-1326;1331-1341;1343-1
A:Cross-references: EMBL:L03710
R:Martindale, D.W.; Taylor, F.M.
Nucleic Acids Res. 16, 2189-2201, 1988
A:Title: Multiple introns in a conjugation-specific gene from Tetrahymena thermophila.
A:Reference number: S03650; MUID:88189811; PMID:3357771
A:Accession: S03650
A:Molecule type: DNA
A:Residues: 236-250,'I',252-255,'N',257-773 <MAR>
A:Cross-references: EMBL:X06462
C:Genetics:
A:Gene: cniB
A:Genetic code: SGC5
A:Introns: 85/3; 136/1; 157/3; 201/2; 290/2; 327/3; 499/1; 573/2; 607/3; 708/3; 777/3; 8
C:Keywords: zinc finger
F:1164-1450/Region: glycine-rich
F:1451-1464/Region: zinc finger CCHC motif
F:1478-1491/Region: zinc finger CCHC motif
F:1501-1514/Region: zinc finger CCHC motif
F:1530-1543/Region: zinc finger CCHC motif
F:1555-1568/Region: zinc finger CCHC motif
F:1579-1592/Region: zinc finger CCHC motif
F:1602-1615/Region: zinc finger CCHC motif
F:1626-1748/Region: glycine-rich

Query Match 12.3%; Score 95; DB 2; Length 1748;
Best Local Similarity 32.0%; Pred. No. 5.6;
Matches 31; Conservative 11; Mismatches 29; Indels 26; Gaps 4;
A:Map position: 2

QY 25 QWGGGNGHNGA---GNSGGPDTLSIYQGSANAALQSDARKSETIT---QSGYGN 77
DB 1640 QFGGGSNGGQSGNGTSSGSDWN-----CQSNVQESTTSSGGWSSSGN 1685

QY 78 GADVGQAGADNLTQNGFRNNATIDQNAKNSDIT 114
DB 1686 QTGGGWSNDN-----QQQWNTGGGWSNSQNT 1717

RESULT 13
E95965
Hypothetical glycine-rich protein [imported] - Sinorhizobium meliloti (strain 1021) mega
C:Species: Sinorhizobium meliloti
C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C:Accession: E95965
R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A:Title: The complete sequence of the 1,683-Rb pSymB megaplasmid from the N2-fixing endo
A:Reference number: A95842; MUID:21396508; PMID:11481431
A:Accession: E95965
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2174 <KUR>
A:Cross-references: GB:A591985; PIDN:CA49389.1; PID:g15140875; GSPDB:GN00167
R:Experimental source: strain 1021, megaplasmid pSymB
R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, P.; Barloy-Hubler,
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A:Reference number: A96039; MUID:21369234; PMID:11474104
A:Contents: annotation
C:Genetics:

A:Gene: SMD21548
A:Genome: plasmid

Query Match 12.0%; Score 93; DB 2; Length 2174;
Best Local Similarity 24.5%; Pred. No. 11;
Matches 36; Conservative 15; Mismatches 32; Indels 64; Gaps 7;
A:Map position: 2

QY 11 AIVVSGSALAGVVPO--WGGGNGHNGGNSGPDSTLSIYQGSANAALQSDARKSET 68
DB 693 AATAGAGAVGILASIGGGG---GGNATGGAGFGSFGGGGGG----- 737
QY 69 TITQSGYNGADVG--QGADNSTIELTQNGFRNNATIDQNAKNSDITVQYDQLVTRVVT 127
DB 738 ---GGYANTANVFGK-----LTLTQ----- 757
QY 128 HEMAHASVMVRQV---GFGNATANQY 151
DB 758 ---SHAAAGIVAQSVGGGGTGGTASSY 781

RESULT 14
B84533
Hypothetical protein At2g15770 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: B84533
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.B.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: B84533
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-301 <STO>
A:Cross-references: GB:AE002093; NID:G5306254; PIDN:AA41987.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g15770
A:Map position: 2

Query Match 11.7%; Score 91; DB 2; Length 301;
Best Local Similarity 26.4%; Pred. No. 1.7;
Matches 33; Conservative 19; Mismatches 51; Indels 22; Gaps 5;
A:Map position: 2

QY 15 SGSALA-----GVVPQWGGGNGH--NGGNSGPDSTLSIYQGSANAALQSDARKSET 68
DB 69 SGGGLGSNNNGSGWGWGTGSGSGSSTNPDCSRSSNNW-----SLKSGWSWSWG 121

QY 69 TITQSGYNGADVGQCADNST---IELTQNGFRNNATIDQNAKNSDITVQYD 119
DB 122 SNDNDSNNSGSDSGGLDRETPKTIIVGGSDGWKKGLDYKDWSKNAPFYVNDLVFKYD 181

QY 120 OLVTTR 124
DB 182 KSAKR 186

RESULT 15
A56038
DNA-binding protein ovo - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 01-Dec-1995 #sequence_revision 01-Dec-1995 #text_change 21-Jul-2000
C:Accession: A56038
R:Garfinkel, M.D.; Wang, J.; Liang, Y.; Mahowald, A.P.
Mol. Cell. Biol. 14, 6809-6818, 1994
A:Title: Multiple products from the shavenbaby-ovo gene region of Drosophila melanogast
A:Reference number: A56038; MUID:95021209; PMID:7935398
A:Accession: A56038
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1028 <GAR>
A:Cross-references: GB:U11383; NID:G520526; PIDN:AAB60216.1; PID:G520527

```

C;Genetics:
A;Gene: ovo
A;Cross-references: FlyBase:FBgn0003028

Query Match      11.7%; Score 91; DB 2; Length 1028;
Best Local Similarity 31.2%; Pred. No. 6.7;
Matches 34; Conservative 8; Mismatches 41; Indels 26; Gaps 4;

QY 3 LKVAFAAIVVSGSALAGVFPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQSD 62
    ||| | : ||| ||| : ||| ||| : ||| ||| : ||| ||| : ||| ||| : |||
Db 59 LQNAAAAAYIMSAGSG-----GGCTGNGGGGAGSGGGPSANSGGGGGG----- 104

QY 63 ARKSETTITQSGYNGADYCGGADNSTIELTONGFRNNATIDOWNAKNS 111
    : ||| ||| ||| : ||| ||| : ||| ||| : ||| ||| : ||| ||| : |||
Db 105 -----CGGYNCGGVG-GPNNNS---LDGNLLNFASVSNYNESNS 141

Search completed: August 2, 2004, 14:56:21
Job time : 9.4 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:36:12 ; Search time 5.3 Seconds
(without alignments)

1483.508 Million cell updates/sec

Title: US-09-543-407-14

Perfect score: 775

Sequence: 1 MKLLKVAFAAIVVSGSALA.....HASMVRQVGFNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	693	89.4	151	1	CSGA_SALTY
2	528	68.1	151	1	CSGA_ECOLI
3	506.5	65.4	152	1	CSGA_BCO57
4	104.5	13.5	151	1	CSGB_ECOLI
5	98.5	12.7	151	1	CSGB_SALTY
6	92	11.9	1327	1	TNKL_HUMAN
7	91	11.7	1028	1	OVO_DROME
8	90.5	11.7	151	1	CSGB_SALTY
9	90	11.6	1093	1	PER_DROWI
10	89.5	11.5	262	1	VG38_BPT2
11	88.5	11.4	479	1	PRTC_ERWCH
12	87	11.2	646	1	GP63_LEIME
13	87	11.2	678	1	YF48_MYCTU
14	86	11.1	599	1	GP63_LEICH
15	85.5	11.0	1048	1	AG01_ARATH
16	85	11.0	347	1	MSA2_PLAF2
17	84	10.8	590	1	GP63_LEIDO
18	84	10.8	602	1	GP63_LEIMA
19	83.5	10.8	1400	1	PD46_CABEL
20	83.5	10.8	1067	1	SGG_DROME
21	83	10.7	362	1	P35_MYCPE
22	83	10.7	1656	1	OMP8_RICJA
23	82	10.6	1080	1	HDC_DROME
24	81.5	10.5	392	1	HMEI_HUMAN
25	81	10.5	165	1	GRP1_ORYSA
26	81	10.5	1612	1	RRFO_PMWUJ
27	81	10.5	1612	1	RRFO_PMWVS
28	80.5	10.4	481	1	PRTB_ERWCH
29	80.5	10.4	641	1	IMD_ARTGO
30	80	10.3	385	1	PER_DROME
31	80	10.3	491	1	YK38_MYCTU
32	80	10.3	548	1	CEAK_ECOLI
33	79.5	10.3	204	1	HEVE_HEVBR

ALIGNMENTS

RESULT 1

ID	CSGA_SALTY	STANDARD	PRT	151 AA
AC	P52257			
DT	01-OCT-1996	(Rel. 34, Last sequence update)		
DT	01-OCT-1996	(Rel. 34, Last sequence update)		
DT	10-OCT-2003	(Rel. 42, Last annotation update)		
DE	Major curlin subunit precursor (Fimbrin SEFI7).			
GN	CSGA OR AGPA OR STM144 OR STY1181 OR T1776.			
OS	Salmonella typhimurium.			
OS	Salmonella typhi, and			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OC	Enterobacteriaceae; Salmonella.			
CX	NCBI_TaxID=602, 601, 592;			
LN	[1]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=S.typhimurium; STRAIN=SR-11;			
RX	MEDLINE=98117058; PubMed=9457880;			
RA	Romling U., Bian Z., Hammar M., Sierralta W.D., Normark S.;			
RT	"Curli fibers are highly conserved between Salmonella typhimurium and			
RT	Escherichia coli with respect to operon structure and regulation.";			
RL	J. Bacteriol. 180:722-731(1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;			
RX	MEDLINE=21534948; PubMed=11677609;			
RA	McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,			
RA	Courtnay L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,			
RA	Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,			
RA	Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,			
RA	Waterston R., Wilson R.K.;			
RT	"Complete genome sequence of Salmonella enterica serovar Typhimurium			
RT	LT2.";			
RL	Nature 413:852-856(2001).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=S.typhi; STRAIN=CT18;			
RX	MEDLINE=21534947; PubMed=11677608;			
RA	Parkhill J., Dougan G., James K.D., Holden M.T.G., Sebatina M.,			
RA	Churcher C., Mungall K.L., Bentley K., Chillingworth T., Connor P.,			
RA	Baker S., Basham D., Brooks K., Davies R.M., Dowd L., White N., Farrar J.,			
RA	Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,			
RA	Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,			
RA	Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Barry C.,			
RA	Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,			
RA	Whitehead S., Barrall B.G.;			
RT	"Complete genome sequence of a multiple drug resistant salmonella			
RT	enterica serovar Typhi CT18.";			
RL	Nature 413:848-852(2001).			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=S.typhi; STRAIN=Ty2 / ATCC 700931;			
RX	MEDLINE=22531367; PubMed=12644504;			
RA	Deng W., Lion S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,			
RA	Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;			

P91698 drosophila
Q8x1f0 homo sapien
P18127 xanthomonas
P30688 neisseria m
Q24180 drosophila
Q04893 saccharomyc
P34291 caenorhabdi
P34308 caenorhabdi
P52302 drosophila
Q12756 homo sapien
P06997 escherichia
P55004 pharbitis n

34 79 10.2 396 1 PER_DROPV
35 79 10.2 933 1 NPA3_HUMAN
36 78.5 10.1 1567 1 ICEN_XANCT
37 78 10.1 331 1 OMB2_NEIMB
38 78 10.1 576 1 DEAF_DROME
39 77.5 10.0 1140 1 YM96_YEAST
40 77 9.9 401 1 YK03_CABEL
41 77 9.9 783 1 YK02_CABEL
42 77 9.9 796 1 MBN_DROME
43 77 9.9 1690 1 KFLA_HUMAN
44 76.5 9.9 329 1 SUBI_ECOLI
45 76.5 9.9 1115 1 PHVE_PHANI

"Comparative genomics of *Salmonella enterica* serovar Typhi strains Ty2 and CT18.";
 J. Bacteriol. 185:2330-2337(2003).
 [5] SEQUENCE FROM N.A.
 RC SPECIES-S. enteritidis; STRAIN=27655-3B;
 RX MEDLINE=96146512; PubMed=8550497;
 RA Collinson S.K., Clouthier S.C., Doran J.L., Baner P.A., Kay W.W.;
 "Salmonella enteritidis agfBAC operon encoding thin, aggregative fimbriae";
 J. Bacteriol. 178:662-667(1996).
 [6] SEQUENCE OF 21-151 FROM N.A.
 RC SPECIES-S. enteritidis; STRAIN=27655-3B;
 RX MEDLINE=96146512; PubMed=8550497;
 RA Doran J.L., Collinson S.K., Burian J., Peterkin P.I., Kay W.W.;
 Munro C.K., Kay C.M., Baner P.A.;
 "DNA-based diagnostic tests for *Salmonella* species targeting agfA, the structural gene for thin, aggregative fimbriae";
 J. Clin. Microbiol. 31:2263-2273(1993).
 [7] SEQUENCE OF 21-33.
 RC SPECIES-S. enteritidis; STRAIN=27655-3B;
 RX MEDLINE=91310586; PubMed=1677357;
 RA Collinson S.K., Emedy L., Mueller K.-M., Trust T.J., Kay W.W.;
 "Purification and characterization of thin, aggregative fimbriae from *Salmonella enteritidis*";
 J. Bacteriol. 173:4773-4781(1991).
 CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO FIBRONECTIN.
 CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 DR EMBL; AJ002301; CAA05317.1; -;
 DR EMBL; AF008749; AAL20074.1; -;
 DR EMBL; AL627269; CAD08268.1; -;
 DR EMBL; AF016840; AAO69399.1; -;
 DR EMBL; U43280; AAC43599.1; -;
 DR PIR; JC6039; JC6039.
 DR StyGene; SGI0608; csGA.
 KW Fimbria; Signal; Complete proteome.
 FT SIGNAL 1 20
 FT CHAIN 21 151 MAJOR CURLIN SUBUNIT.
 FT CONFLICT 134 151 SVMVQVGFGNATANYQ -> DSVTQVAS (IN REF. 6).
 FT SEQUENCE 151 AA; 15305 MW; B7DAC0D16B621359 CRC64;
 Query Match 89.4%; Score 693; DB 1; Length 151;
 Best Local Similarity 90.7%; Pred. No. 4.9e-53;
 Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAALIVSGSALAGVVPWGGGNGHNGGSGPDSTLSIVQYGSANAALQ 60
 DB 1 MKLLKVAFAALIVSGSALAGVVPWGGGNGHNGGSGPDSTLSIVQYGSANAALQ 60
 QY 61 SPARKSETTITOSGVNGADVGQADNSTIETQNGFRNATIDQWNAKNSDITVGYDQ 120
 DB 61 SPARKSETTITOSGVNGADVGQADNSTIETQNGFRNATIDQWNAKNSDITVGYDQ 120
 QY 121 LVTRVVTHEMAHASVWVQVGFGNATANYQ 151
 DB 121 LVTALVNOTASDSSVWVQVGFGNATANYQ 151

RESULT 2
 CSGA_ECOLI STANDARD; PRT; 151 AA.
 ID CSGA_ECOLI
 AC P28307;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Major curlin subunit precursor.
 GN CSGA OR B1042.
 OS *Escherichia coli*.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; *Escherichia*.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / W3110;
 RX MEDLINE=93211294; PubMed=8459772;
 RA Olsen A., Arngqvist A.;
 "The RpoS sigma factor relieves H-NS-mediated transcriptional repression of csGA, the subunit gene of fibronectin-binding curli in *Escherichia coli*";
 Mol. Microbiol. 7:523-536(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MC4100;
 RX MEDLINE=96414468; PubMed=8817489;
 RA Hammar M., Arngqvist A., Bian Z., Olsen A., Normark S.;
 "Expression of two csG operons is required for production of fibronectin- and congo red-binding curli polymers in *Escherichia coli* K-12";
 Mol. Microbiol. 18:661-670(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V., Riley M., Collado-VIDES J., Glasner J.D., Rode C.K., Mayhew G.F., Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J., Mau B., Shao Y.;
 "The complete genome sequence of *Escherichia coli* K-12";
 Science 277:1453-1474(1997).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97061202; PubMed=8905232;
 RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A., Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K., Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K., Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N., Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y., Yano M., Horiuchi T.;
 "A 718-kb DNA sequence of the *Escherichia coli* K-12 genome corresponding to the 12.7-28.0 min region on the linkage map";
 DNA Res. 3:137-155(1996).
 RN [5]
 RP SEQUENCE OF 21-40.
 RC STRAIN=K12 / YMEL;
 RX MEDLINE=93023873; PubMed=1357528;
 RA Arngqvist A., Olsen A., Pfeifer J., Russell D.G., Normark S.;
 "The Crl protein activates cryptic genes for curli formation and fibronectin binding in *Escherichia coli* HB101";
 Mol. Microbiol. 6:2443-2452(1992).
 RN [6]
 RP SEQUENCE OF 21-31.
 RX MEDLINE=91310586; PubMed=1677357;
 RA Collinson S.K., Emedy L., Trust T.J., Kay W.W.;
 "Purification and characterization of thin, aggregative fimbriae from *Salmonella enteritidis*";
 J. Bacteriol. 173:4773-4781(1991).
 CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO FIBRONECTIN.

```
CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L04979; AAA23616.1; -
DR EMBL; X90754; CAA62282.1; -
DR EMBL; AE000205; AAC74126.1; -
DR EMBL; D90741; BAA35832.1; -
DR EMBL; D90742; BAA35840.1; -
DR EMBL; D90788; S70788.
DR EcoGene; EG11489; csGA.
KW Fimbria; Signal; Complete proteome.
FT SIGNAL 1 20 MAJOR CURLIN SUBUNIT.
FT CHAIN 7 151
FT CONFLICT 7 7 A -> E (IN REF. 1).
SQ SEQUENCE 151 AA; 15049 MW; C003470D208D395F CRC64;

Query Match 68.1%; Score 528; DB 1; Length 151;
Best Local Similarity 68.9%; Pred. No. 7,9e-39;
Matches 104; Conservative 20; Mismatches 27; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGHHNGGNSGDPSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAATAAIVFSSALAGVVPQYGGGHHNGGNSGNSSEINLIYQYGGNSALALQ 60
QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGVQYDQ 120
DB 61 TDARNSDLITITQHGCGNGADVGQSDSSIDLTRQFGNSATLDQWNGKNSMTVVKQFG 120

QY 121 LVTRVVTHEMAHASVMVQVGFNNATANQY 151
DB 121 GNCGAAVDQTASNSVNVTVQVGFNNATAHQY 151

RESULT 3
CSGA_ECO57 STANDARD; PRT; 152 AA.
ID CSGA_ECO57
AC Q93U24;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Major curlin subunit precursor.
GN CSGA OR Z1676 OR ECS1420.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]_TaxID=83334;
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / ATCC 43895;
RX MEDLINE=21218556; PubMed=11206551;
RA Perna N.T., Plunkett G., III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
RT Nature 409:529-533(2001).
```

```
[3]
RN RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Havaishi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kunara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12."
RL DNA Res. 8:11-22(2001).
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN.
CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF275733; AAK53212.1; -
DR EMBL; AE005315; AAG55788.1; -
DR EMBL; AP002554; BAB34843.1; -
DR PIR; D90806; D90806.
DR PIR; H85665; H85665.
KW Fimbria; Signal; Complete proteome.
FT SIGNAL 1 20 BY SIMILARITY.
FT CHAIN 21 152 MAJOR CURLIN SUBUNIT.
FT SEQUENCE 152 AA; 15099 MW; EE2D2D94DDE91243 CRC64;

Query Match 65.4%; Score 506.5; DB 1; Length 152;
Best Local Similarity 67.1%; Pred. No. 5,6e-37;
Matches 102; Conservative 21; Mismatches 28; Indels 1; Gaps 1;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGHHNGGNSGDPSTLSIYQYGSANAALAL 59
DB 1 MKLLKVAATAAIVFSSALAGVVPQYGGGHHNGGNSGNSSEINLIYQYGGNSALAL 60
QY 60 QSDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGVQYD 119
DB 61 QADARNSDLITITQHGCGNGADVGQSDSSIDLTRQFGNSATLDQWNGKNSMTVVKQFG 120
QY 120 QLVTRVVTHEMAHASVMVQVGFNNATANQY 151
DB 121 GNCGAAVDQTASNSVNVTVQVGFNNATAHQY 152

RESULT 4
CSGB_ECOLI STANDARD; PRT; 151 AA.
ID CSGB_ECOLI
AC P39828;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Minor curlin subunit precursor.
GN CSGB OR B1041 OR Z1675 OR ECS1419.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]_TaxID=562, 83334;
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MC4100;
RX MEDLINE=96414468; PubMed=8817489;
RA Hammar M., Arngqvist A., Bian Z., Olsen A., Normark S.;
RT "Expression of two csG operons is required for production of
RT fibronectin- and Congo red-binding curli polymers in Escherichia coli
```

RT K-12.";

RL Mol. Microbiol. 18:661-670(1995).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=K12 / MG1655;

RX MEDLINE=97426617; PubMed=9278503;

RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V., Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F., Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J., Mau B., Shao Y.;

RA "The complete genome sequence of *Escherichia coli* K-12.";

RL Science 277:1453-1474(1997).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=K12;

RX MEDLINE=97061202; PubMed=8905232;

RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A., Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K., Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K., Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N., Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y., Yano M., Horiuchi T.;

RA "A 718-kb DNA sequence of the *Escherichia coli* K-12 genome corresponding to the 12.7-28.0 min region on the linkage map.";

RL DNA Res. 3:137-155(1996).

RN [4]

RP SEQUENCE FROM N.A.

RC STRAIN=O157:H7 / EDL933 / ATCC 700927;

RX MEDLINE=21074935; PubMed=11206551;

RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D., Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A., Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L., Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamocousis K., Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C., Welch R.A., Blattner F.R.;

RA "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7.";

RL Nature 409:529-533(2001).

RN [5]

RP SEQUENCE FROM N.A.

RC STRAIN=O157:H7 / RIMD 0509952;

RX MEDLINE=21156231; PubMed=11258796;

RA Hayaishi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K., Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T., Tada T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T., Kuhara S., Shiba T., Hattori M., Shinagawa H.;

RA "Complete genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12.";

RL DNA Res. 8:11-22(2001).

RN [6]

RP SEQUENCE OF 1-21 FROM N.A.

RC STRAIN=K12;

RX MEDLINE=95157246; PubMed=7854117;

RA Arngqvist A., Olsen A., Normark S.;

RA "Sigma S-dependent growth-phase induction of the *csgBA* promoter in *Escherichia coli* can be achieved in vivo by sigma 70 in the absence of the nucleoid-associated protein H-NS.";

RL Mol. Microbiol. 13:1021-1032(1994).

CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF CURLIN MONOMERS.

CC -!- SIMILARITY: BELONGS TO THE CSGB/CSGB FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

CC EMBL; X90754; CAA62281.1; -.

DR EMBL; AE000205; AAC74125.1; -.

DR EMBL; D90741; BAA35831.1; -.

DR EMBL; AE005315; AAG55787.1; -.

DR EMBL; AP002554; BAB34842.1; -.

DR PIR; C90806; C90806.

DR PIR; G85665; G85665.

DR PIR; S70787; S70787.

DR EcoGene; EGI2621; csGB.

KW Fimbria; Signal; Complete proteome.

FT SIGNAL 1 21 POTENTIAL.

FT CHAIN 22 151 MINOR CURLIN SUBUNIT.

SQ SEQUENCE 151 AA; 15882 MW; B18D26B96401488 CRC64;

Query Match 13.5%; Score 104.5; DB 1; Length 151;

Best Local Similarity 29.2%; Pred No. 0.023;

Matches 35; Conservative 19; Mismatches 59; Indels 7; Gaps 4;

Qy 9 FAATVWGSAL--AGVVPQWGGNGNNGSSGPDSTLSIY-QYGSANAALALQSDARK 65

Db 33 FAVNELSKSFNQAAII--GQAGTNNSAQLRQGGSKLLAVVAQEGSSNRA-KIDQTGDY 88

Qy 66 SETTITSGYNGADYCGAGDNSTIELTQNGFENNATIDOWNAKNSDITVGQYDQLVTRV 125

Db 89 NLAYIDAGSANDASISQAGYNTAMIIQKSGNKANITQYGTQKTAIVVQSQVAIRV 148

RESULT 5

CSGB_SALTY

ID_CSGB_SALTY STANDARD; PRT; 151 AA.

AC P55225;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Minor curlin subunit precursor (Fimbrin SEF17 minor subunit).

GN CSGB OR AGFB OR STM1143.

OS *Salmonella typhimurium*, and

OS *Salmonella enteritidis*.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; *Salmonella*.

OX NCBI_TaxID=602, 592;

RN [1]

RP SEQUENCE FROM N.A.

RC SPECIES=S.typhimurium; STRAIN=SE-11;

RX MEDLINE=98117058; PubMed=9457880;

RA Rmiling U., Bian Z., Hammar M., Sierralta W.D., Normark S.;

RA "Curli fibers are highly conserved between *Salmonella typhimurium* and *Escherichia coli* with respect to operon structure and regulation.";

RL J. Bacteriol. 180:722-731(1998).

RN [2]

RP SEQUENCE FROM N.A.

RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;

RX MEDLINE=21534948; PubMed=11677609;

RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P., Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D., Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E., Ryan B., Sun H., Florea L., Miller W., Stoneking T., Nhan M., Waterston R., Wilson R.K.;

RA "Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2.";

RL Nature 413:852-856(2001).

RN [3]

RP SEQUENCE FROM N.A.

RC SPECIES=S.enteritidis; STRAIN=27655-3B;

RX MEDLINE=96146512; PubMed=8550497;

RA Collinson S.K., Clouthier S.C., Doran J.L., Bansen P.A., Kay W.W.;

RA "Salmonella enteritidis *agfBAC* operon encoding thin, aggregative fimbriae.";

RL J. Bacteriol. 178:662-667(1996).

CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF CURLIN MONOMERS.

CC -----

CC EMBL; X90754; CAA62281.1; -.


```

CC  -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ002301; CAA05316.1; -.
DR EMBL; AE008749; AAL20073.1; -.
DR EMBL; U63280; AAC43598.1; -.
DR PIR; JC6040; JC6040.
DR StyGene; SG10609; csGB.
KW Fibria; Signal; Complete proteome.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 151 MINOR CURLIN SUBUNIT.
SQ SEQUENCE 151 AA; 16182 MW; CQFC5430B6DD361D CRC64;
Query Match 12.7%; Score 98.5; DB 1; Length 151;
Best Local Similarity 26.2%; Pred. No. 0.074;
Matches 34; Conservative 17; Mismatches 42; Indels 37; Gaps 4;
QY 10 AAIIVVGSALGVVPQGGGNGHNGNSGPDSTLSIYQVGSANAALQSDARKSETT 69
D5 58 ARVROEGSKLLSVISQ--EGGNRAKVQAGNFAYIEQTGNAN-----DAS 103
QY 70 ITQSGYNGADVGQGDNSTIELQNGFRNATIDWNKNSDITVGQDQVTRVVTHE 129
D5 104 ISQAYGNNA-----AIQKSGNKANTIQGTQK-----TAVVQVK 140
QY 130 MAHASVMVRQ 139
D5 141 QSHVAIRVTQ 150
RESULT 6
ID TNK1 HUMAN STANDARD; PRT; 1327 AA.
AC Q95271; Q95272;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Tankyrase 1 (EC 2.4.2.30) (TANK1) (Tankyrase I) (TNKS-1) (TRF1-
DE interacting ankyrin-related ADP-ribose polymerase).
GN TNKS OR TNKS1 OR TIN1 OR TIN1 OR PARPL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC TISSUE=Testis;
RX MEDLINE=99040105; PubMed=9822378;
RA Smith S., Giriat L, Schmitt A., de Lange T.;
RT "Tankyrase, a poly(ADP-ribose) polymerase at human telomeres.";
RL Science 282:1484-1487(1998).
RN [2]
RP SUBCELLULAR LOCATION.
RX MEDLINE=99454782; PubMed=10523501;
RA Smith S., de Lange T.;
RT "Cell cycle dependent localization of the telomeric PARP, tankyrase,
RT to nuclear pore complexes and centrosomes.";
RL J. Cell Sci. 112:3649-3656(1999).
RN [3]
RP FUNCTION, AND PHOSPHORYLATION.
RX MEDLINE=20556282; PubMed=10988299;
RA Chi N.-W., Iodish H.F.;
RT "Tankyrase is a Golgi-associated mitogen-activated protein kinase
RT substrate that interacts with IRAP in GLUT4 vesicles.";
RL J. Biol. Chem. 275:38437-38444(2000).
RN [4]

```

```

RP FUNCTION, AND MUTAGENESIS OF HIS-1184 AND GLU-1291.
RX MEDLINE=21602874; PubMed=11739745;
RA Cook B.D., Dynsek J.N., Chang W., Shostak G., Smith S.;
RT "Role for the related poly(ADP-ribose) polymerases tankyrase 1 and 2
RT at human telomeres.";
RL Mol. Cell. Biol. 22:332-342(2002).
CC -!- FUNCTION: May regulate vesicle trafficking and modulate the
CC subcellular distribution of SLC2A4/GLUT4-vesicles. Has PARP
CC activity and can modify TRF1, and thereby contribute to the
CC regulation of telomere length.
CC -!- CATALYTIC ACTIVITY: NAD(+) + [ADP-D-riboseyl](N)-acceptor =
CC nicotinamide + {ADP-D-riboseyl}(N+1)-acceptor.
CC -!- SUBUNIT: Oligomerizes and associates with TNKS2. Interacts with
CC the cytoplasmic domain of INPEP/Otase in SLC2A4/GLUT4-vesicles.
CC Binds to the N-terminus of telomeric TRF1 via the ANK repeats.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic, associated with the Golgi and
CC with juxtanuclear SLC2A4/GLUT4-vesicles. A minor proportion is
CC also found at nuclear pore complexes and around the pericentriolar
CC matrix of mitotic centrosomes. During interphase, a small fraction
CC of TNKS is found in the nucleus, associated with TRF1.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=O95271-1; Sequence=Displayed;
CC Name=2;
CC IsoId=O95271-2; Sequence=VSP_004538, VSP_004539;
CC Note=No experimental confirmation available;
CC -!- TISSUE SPECIFICITY: Ubiquitous; highest levels in testis.
CC -!- PTM: Upon insulin-stimulation, phosphorylated on serine residues
CC by VAPK kinases.
CC -!- PTM: ADP-ribosylated (-auto).
CC -!- SIMILARITY: Belongs to the PARP family.
CC -!- SIMILARITY: Contains 15 ANK repeats.
CC -!- SIMILARITY: Contains 1 sterile alpha motif (SAM) domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF082556; AAC79841.1; -.
DR EMBL; AF082557; AAC79842.1; -.
DR EMBL; AF082558; AAC79843.1; -.
DR EMBL; AF082559; AAC79844.1; -.
DR HSSP; Q00420; LAWC.
DR Genew; HGNC:11941; TNKS.
DR MIM; 603303; -.
DR GO; GO:0000781; C:chromosome, telomeric region; IDA.
DR GO; GO:0003950; F:NAD ADP-ribosyltransferase activity; IDA.
DR GO; GO:0005515; F:protein binding; IPI.
DR GO; GO:0007004; P:telomerase-dependent telomere maintenance; NAS.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR001660; SAM.
DR Pfam; PF00023; ank; 13.
DR Pfam; PF00536; SAM; 1.
DR PRINTS; PR01415; ANKYRIN.
DR SMART; SM00248; ANK; 17.
DR SMART; SM00454; SAM; 1.
DR PROSITE; PS50088; ANK_REPEAT; 15.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS50105; SAM_DOMAIN; 1.
KW Transferase; Glycosyltransferase; NAD; Golgi stack; Telomere;
KW Nuclear protein; Repeat; ANK repeat; ADP-ribosylation;
KW Phosphorylation; Alternative splicing.
FT REPEAT 215 247 ANK 1.
FT REPEAT 248 280 ANK 2.
FT REPEAT 281 313 ANK 3.
FT REPEAT 368 400 ANK 4.
FT REPEAT 401 433 ANK 5.
FT REPEAT 434 466 ANK 6.

```

```

FT REPEAT 521 556 ANK 7.
FT REPEAT 557 589 ANK 8.
FT REPEAT 590 622 ANK 9.
FT REPEAT 683 715 ANK 10.
FT REPEAT 716 748 ANK 11.
FT REPEAT 749 781 ANK 12.
FT REPEAT 836 868 ANK 13.
FT REPEAT 869 901 ANK 14.
FT REPEAT 902 934 ANK 15.
FT DOMAIN 1030 1089 SAM.
FT DOMAIN 1176 1327 PARP.
FT DOMAIN 9 14 POLY-HIS.
FT DOMAIN 27 34 POLY-PRO.
FT DOMAIN 128 134 POLY-SER.
FT DOMAIN 137 145 POLY-SER.
FT VARSPLIC 641 643 EST -> GHS (in isoform 2).
FT VARSPLIC 644 1327 /FTID=VSP 004538.
FT VARSPLIC 644 1327 Missing (in isoform 2).
FT MUTAGEN 1184 1184 /FTID=VSP 004539.
FT MUTAGEN 1291 1291 H->A: LOSS OF ACTIVITY; WHEN ASSOCIATED
FT MUTAGEN 1291 1291 E->A: LOSS OF ACTIVITY; WHEN ASSOCIATED
FT SEQUENCE 1327 AA; E14DE985C710B957 CRC64;

Query Match 11.9%; Score 92; DB 1; Length 1327;
Best Local Similarity 30.4%; Pred. No. 3.3;
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

QY 6 VAAFAAI-VVSGSALAGVFPQWGGGNGHNGSGSPDSTLSIYQYGSANAALQSDAR 64
Db 99 VAAAPVPAVTSRAAGVAPNAGSGSNNPSSSSPTSS-SSSSPSPSGSLASPEAA 157
QY 65 KSETTT-----OSGYNGADVCGQADNSTIETQNG--FRNNTIDQWNAKNSDI 113
Db 158 GVSSTAPLFGGAAGPCTGPAVSGALRELLACRNGDVSRVKRLVDAANVAKDM 212

RESULT 7
OVO_DROME STANDARD; PRT; 1028 AA.
AC P51521; Q9XZU4;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ovo protein (Shaven baby protein).
GN OVO OR SVB.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovary;
RX MEDLINE=95021209; PubMed=7935398;
RA Garfinkel M.D., Wang J., Liang Y., Mahowald A.P.;
RT "Multiple products from the shavenbaby-ovo gene region of Drosophila
RT melanogaster: relationship to genetic complexity.";
RL Mol. Cell. Biol. 14:6809-6818(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Oregon-R;
RX MEDLINE=91293102; PubMed=1712294;
RA Mevel-Ninio M.T.M., Terracol R., Kafatos F.C.;
RT "The ovo gene of Drosophila encodes a zinc finger protein required
RT for female germ line development.";
RL EMBO J. 10:2259-2266(1991).
CC -!- FUNCTION: REQUIRED FOR SURVIVAL AND DIFFERENTIATION OF FEMALE GERM
CC LINE CELLS. PLAYS A ROLE IN GERM LINE SEX DETERMINATION.
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -!- DEVELOPMENTAL STAGE: FIRST APPEARS IN THE GERMARIIUM AND
CC ACCUMULATES IN NURSE CELLS DURING OOGENESIS. STORED IN THE EGG,

```

```

CC BUT IS RAPIDLY LOST IN THE EMBRYOS EXCEPT FOR ITS CONTINUED
CC PRESENCE IN THE GERM LINE PRECURSOR POLE CELLS.
CC -!- SIMILARITY: Contains 4 C2H2-type zinc fingers.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; U11383; AB60216.1; -.
CC EMBL; X59772; CAB36921.1; ALT_SEQ.
CC PIR; A56038; A56038.
CC HSP; P07248; 2ADR.
CC TRANSFAC; T00669; -.
CC FlyBase; FBgn003028; ovo.
CC InterPro; IPR007087; Znf_C2H2.
CC Pfam; PF00036; zf-C2H2; 3.
CC SMART; SM00355; Znf_C2H2; 4.
CC PROSITE; PS00028; ZINC FINGER C2H2_1; 3.
CC PROSITE; PS0157; ZINC FINGER C2H2_2; 3.
CC Zinc-finger; Metal-binding; DNA-binding; Repeat; Nuclear protein;
KW Transcription regulation.
FT DOMAIN 62 66 POLY-ALA.
FT DOMAIN 72 77 POLY-GLY.
FT DOMAIN 80 85 POLY-GLY.
FT DOMAIN 98 108 POLY-HIS.
FT DOMAIN 144 152 POLY-ASN.
FT DOMAIN 133 159 POLY-GLN.
FT DOMAIN 336 339 POLY-GLN.
FT DOMAIN 347 353 POLY-GLN.
FT DOMAIN 357 361 POLY-GLN.
FT DOMAIN 410 414 POLY-GLN.
FT DOMAIN 418 422 POLY-GLN.
FT DOMAIN 426 432 POLY-GLN.
FT DOMAIN 445 453 POLY-GLN.
FT DOMAIN 456 459 POLY-GLN.
FT DOMAIN 466 474 POLY-GLN.
FT DOMAIN 497 517 POLY-ALA.
FT DOMAIN 524 529 POLY-SER.
FT DOMAIN 549 558 POLY-ALA.
FT DOMAIN 639 651 POLY-ALA.
FT DOMAIN 717 725 POLY-ALA.
FT DOMAIN 797 802 POLY-GLN.
FT DOMAIN 820 823 POLY-GLN.
FT DOMAIN 826 832 POLY-GLN.
FT ZN_FING 874 896 C2H2-TYPE 1.
FT ZN_FING 902 924 C2H2-TYPE 2.
FT ZN_FING 930 953 C2H2-TYPE 3.
FT ZN_FING 969 992 C2H2-TYPE 4.
FT CONFLICT 647 647 A -> R (IN REF. 2).
SQ SEQUENCE 1028 AA; 110620 MW; D70688B2BC0F6F77 CRC64;

Query Match. 11.7%; Score 91; DB 1; Length 1028;
Best Local Similarity 31.2%; Pred. No. 3;
Matches 34; Conservative 8; Mismatches 41; Indels 26; Gaps 4;

QY 3 LLKVAFAAIVVSGSALAGVFPQWGGGNGHNGSGSPDSTLSIYQYGSANAALQSD 62
Db 59 LQNAAAAAYIMSAGSG-----GGCTGNGGGGASGPGGFSANSGGGGGGG----- 104
QY 63 ARKSETTITQSGYNGADVCGQADNSTIETQNGFRNNTIDQWNAKNS 111
Db 105 -----GGNGYINGGVG-CPNNS---LDGNLNLINFAVSNNESNS 141

RESULT 8
CSGB_SALTI STANDARD; PRT; 151 AA.
AC Q8Z7M3;
DT 28-FEB-2003 (Rel. 41, Created)

```



```

DR EMBL; U51065; AAB41370.1; -
DR EMBL; U51066; AAB41371.1; -
DR EMBL; U51067; AAB41372.1; -
DR EMBL; U51068; AAB41373.1; -
DR EMBL; U51069; AAB41374.1; -
DR EMBL; U51070; AAB41375.1; -
DR EMBL; U51071; AAB41376.1; -
DR EMBL; U51072; AAB41377.1; -
DR EMBL; L06342; AAA28765.1; -
DR FlyBase; FBcr0013161; Dmil\per.
DR InterPro; IPR000014; PAS_domain.
DR Pfam; PF00989; PAS; 2.
DR SMART; SM00091; PAS; 2.
DR PROSITE; PS0112; PAS; 2.
DR Biological rhythms; Repeat; Nuclear protein; Phosphorylation;
KW Polymorphism.
FT NON_TER 1
FT DOMAIN <1 12 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 139 209 PAS 1.
FT DOMAIN 289 359 PAS 2.
FT DOMAIN 371 411 PAC.
FT DOMAIN 7 12 POLY-LYS.
FT DOMAIN 618 625 POLY-GLY.
FT DOMAIN 718 734 POLY-GLY.
FT DOMAIN 745 748 POLY-SER.
FT DOMAIN 759 770 POLY-GLY.
FT DOMAIN 885 POLY-ALA.
FT DOMAIN 911 917 POLY-ALA.
FT VARIANT 611 611 T -> A (IN STRAIN 0811.4).
FT VARIANT 617 617 S -> F (IN STRAIN 0811.4).
FT VARIANT 622 622 G -> V (IN STRAIN GUANA).
FT VARIANT 724 724 G -> A (IN STRAIN MANAUS 2).
FT VARIANT 726 726 G -> S (IN STRAIN SANTA MARIA).
FT VARIANT 729 734 MISSING (IN STRAIN PORTO ALEGRE 3).
FT VARIANT 730 734 MISSING (IN STRAIN PORTO ALEGRE 4).
FT VARIANT 731 734 MISSING (IN STRAINS MANAUS 1 AND MANAUS 3).
FT VARIANT 732 734 MISSING (IN STRAINS LIMA B, L'HABITATUE AND CANO MORA).
FT VARIANT 733 733 G -> V (IN STRAIN PORTO ALEGRE 4).
FT VARIANT 733 734 MISSING (IN STRAINS GUADELOUPE, MANAUS 2, PORTO ALEGRE 2, PORTO ALEGRE 1 AND GUANA).
FT VARIANT 734 734 MISSING (IN STRAINS MANAUS 4, PORTO ALEGRE 1 AND PORTO ALEGRE 2).
FT VARIANT 747 747 S -> A (IN STRAINS GUADELOUPE AND GUANA).
FT VARIANT 764 766 MISSING (IN STRAIN MANAUS 3).
FT VARIANT 886 886 A -> T (IN STRAIN 0811.4).
FT NON_TER 1093 1093
SQ SEQUENCE 1093 AA; 115896 MW; AB6DR050267EC187 CRC64;

Query Match 11.6%; Score 90; DB 1; Length 1093;
Best Local Similarity 27.7%; Pred. No. 3.9; Indels 16; Gaps 2;
Matches 26; Conservative 6; Mismatches 46;

QY 27 GGGGHHNGGSSGDPSTLSIYQYGSANAALQSDARKSETTITQSGYNGADVGGAD 86
DB 721 GGGGGGGGGGGGLFLDVTHTSS-----SSQNKGPVGAAGGAGGGVGGGG-- 770

QY 87 NSTIELTQGFNNATIDQWNAKNSDITVGYDQ 120
DB 771 -----SCSLGNGNGVSGNGNNSQPSNTQYTQ 798

RESULT 10
VG38_BPT2
ID_VG38_BPT2 STANDARD; PRT; 262 AA.
AC_F07875;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 01-AUG-1990 (Rel. 15, Last annotation update)
DE Receptor recognizing protein (Protein Gp38).
GN 38.

OS Bacteriophage T2.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
OC T4-like viruses.
OX NCBI_TaxID=10664;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87283911; PubMed=3302276;
RA Riede I., Drexler K., Eschbach M.L., Henning U.;
RT "DNA sequence of genes 38 encoding a receptor-recognizing protein of bacteriophages T2, K3 and of K3 host range mutants.";
PL J. Mol. Biol. 194:31-39(1987).
CC -!- FUNCTION: Vg38 is at the tip of the long tail fibers and serves as the phage recognition site for the cellular receptor.
CC -!- MISCELLANEOUS: THIS PHAGE USE OUTER MEMBRANE PROTEINS OMPF AND TTR AS RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X05312; CAA28935.1; -
DR PIR; S00275; S00275; Tail_fibre_GP38.
DR InterPro; IPR007932; Tail_fibre_GP38.
DR Pfam; PF05268; GP38; 1.
DR Fiber protein; Phage recognition.
SQ SEQUENCE 262 AA; 25801 MW; 0567366918F6C745 CRC64;

Query Match 11.5%; Score 89.5; DB 1; Length 262;
Best Local Similarity 34.1%; Pred. No. 0.83; Indels 17; Gaps 4;
Matches 30; Conservative 8; Mismatches 33;

QY 27 GGGGHHNGGSSGDPSTLSIYQYGSANAALQSDARKSETTITQSGYNGADVGGAD 86
DB 175 GGGGPFVGGKIGSDSILS-----GSNASL---TDAGTGGTTF-QYGAGNGNGVGGGG 225

QY 87 NSTIELTQGFNNATIDQWNAKNSDIT 114
DB 226 -----RGWKNVYTSFGAAGAAVT 245

RESULT 11
PRTC_ERWCH
ID_PRTC_ERWCH STANDARD; PRT; 479 AA.
AC_P16317;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Secreted protease C precursor (EC 3.4.24.-) (Proc).
GN PRTC.
OS Erwinia chrysanthemi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pectobacterium.
OX NCBI_TaxID=556;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=B374;
RX MEDLINE=91009140; PubMed=2211614;
RA Delepeleire P., Wandersman C.;
RT "Protein secretion in Gram-negative bacteria. The extracellular metalloprotease B from Erwinia chrysanthemi contains a C-terminal secretion signal analogous to that of Escherichia coli alpha-hemolysin.";
PL J. Biol. Chem. 265:17118-17125(1990).
RN [2]
RP SEQUENCE OF 1-59 FROM N.A., AND PARTIAL SEQUENCE.
RX STRAIN=B374;
RX MEDLINE=89255387; PubMed=2722818;
RA Delepeleire P., Wandersman C.;
RT "Protease secretion by Erwinia chrysanthemi. Proteases B and C are

```

FT synthesized and secreted as zymogens without a signal peptide.";
 RL J. Biol. Chem. 264:9083-9089(1989).
 CC -!- COPACTOR: Binds 1 zinc ion and 7 calcium ions per subunit (By
 CC similarity).
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- DONATION: THE GLY-RICH REPEATS MAY BE IMPORTANT IN THE EXTRACELLULAR
 CC SECRETION OF THIS METALLOPROTEASE.
 CC -!- SIMILARITY: Belongs to peptidase family M10B.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M59229; AAA24860.1; -;
 DR EMBL; J04736; AAA24862.1; -;
 DR EMBL; M60395; AAA63638.1; -;
 DR PIR; A38307; A38307.
 DR PDB; 1GO7; 17-OCT-02.
 DR PDB; 1GO8; 17-OCT-02.
 DR PDB; 1K7G; 19-OCT-02.
 DR PDB; 1K7Q; 20-OCT-02.
 DR MEROPS; M10.054; -;
 DR InterPro; IPR001343; Hemolysin_Ca_bind.
 DR InterPro; IPR006025; Pept_M_Zn_BS.
 DR InterPro; IPR006026; Peptidase_M.
 DR Pfam; PF00353; hemolysinCaBind; 3.
 DR PRINTS; PR00313; CASNDNGRPT.
 DR SMART; SMO0235; ZnMC; 1.
 DR PROSITE; PS00142; ZINC_PROTEASE; 1.
 DR PROSITE; PS00330; HEMOLYSIN_CALCULUM; 1.
 KW Hydrolyase; Metalloprotease; Calcium-binding; Metal-binding; Zinc;
 KW Repeat; Zymogen; 3D-structure.
 FT PROPEP 1 17
 FT CHAIN 18 479 SECRETED PROTEASE C.
 FT REPEAT 345 353 GGXGD 1.
 FT REPEAT 354 362 GGXGD 2.
 FT REPEAT 363 371 GGXGD 3.
 FT REPEAT 372 380 GGXGD 4.
 FT REPEAT 381 389 GGXGD 5.
 FT METAL 188 188 ZINC (CATALYTIC) (BY SIMILARITY).
 FT ACT_SITE 189 189 BY SIMILARITY.
 FT METAL 192 192 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 228 228 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 265 265 CALCULUM 1 (VIA CARBONYL OXYGEN) (BY
 FT SIMILARITY).
 FT METAL 267 267 CALCULUM 1 (VIA CARBONYL OXYGEN) (BY
 FT SIMILARITY).
 FT METAL 297 297 CALCULUM 1 (BY SIMILARITY).
 FT METAL 299 299 CALCULUM 1 (VIA CARBONYL OXYGEN) (BY
 FT SIMILARITY).
 FT METAL 300 300 CALCULUM 2 (VIA CARBONYL OXYGEN) (BY
 FT SIMILARITY).
 FT METAL 302 302 CALCULUM 1 AND 2 (BY SIMILARITY).
 FT METAL 339 339 CALCULUM 2 (VIA CARBONYL OXYGEN) (BY
 FT SIMILARITY).
 FT METAL 341 341 CALCULUM 2 (BY SIMILARITY).
 FT METAL 346 346 CALCULUM 3 (VIA CARBONYL OXYGEN) (BY
 FT SIMILARITY).
 FT METAL 348 348 CALCULUM 3 (VIA CARBONYL OXYGEN) (BY
 FT SIMILARITY).
 FT METAL 350 350 CALCULUM 3 (BY SIMILARITY).
 FT METAL 355 355 CALCULUM 4 (VIA CARBONYL OXYGEN) (BY
 FT SIMILARITY).
 FT METAL 357 357 CALCULUM 4 (VIA CARBONYL OXYGEN) (BY
 FT SIMILARITY).
 FT METAL 359 359 CALCULUM 4 (BY SIMILARITY).
 FT METAL 363 363 CALCULUM 3 (VIA CARBONYL OXYGEN) (BY
 FT SIMILARITY).
 FT METAL 364 364 CALCULUM 5 (VIA CARBONYL OXYGEN) (BY
 FT SIMILARITY).

FT METAL 365 365 SIMILARITY).
 FT METAL 366 366 CALCULUM 3 (VIA CARBONYL OXYGEN) (BY
 FT SIMILARITY).
 FT METAL 368 368 CALCULUM 3 AND 5 (BY SIMILARITY).
 FT METAL 372 372 CALCULUM 4 (VIA CARBONYL OXYGEN) (BY
 FT SIMILARITY).
 FT METAL 373 373 CALCULUM 6 (VIA CARBONYL OXYGEN) (BY
 FT SIMILARITY).
 FT METAL 375 375 CALCULUM 6 (VIA CARBONYL OXYGEN) (BY
 FT SIMILARITY).
 FT METAL 377 377 CALCULUM 4 AND 6 (BY SIMILARITY).
 FT METAL 381 381 CALCULUM 5 (VIA CARBONYL OXYGEN) (BY
 FT SIMILARITY).
 FT METAL 382 382 CALCULUM 7 (VIA CARBONYL OXYGEN) (BY
 FT SIMILARITY).
 FT METAL 383 383 CALCULUM 5 (VIA CARBONYL OXYGEN) (BY
 FT SIMILARITY).
 FT METAL 384 384 CALCULUM 7 (VIA CARBONYL OXYGEN) (BY
 FT SIMILARITY).
 FT METAL 386 386 CALCULUM 5 AND 7 (BY SIMILARITY).
 FT METAL 395 395 CALCULUM 6 (BY SIMILARITY).
 FT METAL 402 402 CALCULUM 6 (BY SIMILARITY).
 FT METAL 412 412 CALCULUM 7 (BY SIMILARITY).
 SQ SEQUENCE 479 AA; 51600 MW; 990ED8376725DF61 CRC64;
 Query Match 11.4%; Score 88.5; DB 1; Length 479;
 Best Local Similarity 28.8%; Pred. No. 2;
 Matches 40; Conservative 21; Mismatches 41; Indels 37; Gaps 9;
 QY 27 GGGNHNGGNSGGPDSLTLSIYQGSANALALQDARKSETITQSG-----YNG----- 78
 DB 345 GSGNDILVNSAD-----NILQGGAGNDLVLYGGAGA---DTLYGGAGRTFTVYSGQDS 396
 QY 79 -----ADVGGAGDNSTIETQNGFRNNATII---DOWNAKNSDITVGYQYD--CLVTR 124
 DB 397 TVAAVDWIADFQKID----KIDLSAFRNEGSLSFVQDQFTGKGQEVML-QWDANSIIN 451
 QY 125 VVTHEVAHASV--MVRQVG 141
 DB 452 LMLHEAGHSVDLVRIVG 470
 RESULT 12
 GP63 LEIME
 ID GP63 LEIME STANDARD; PRT; 646 AA.
 AC P43150;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Leishmanolysin C1 precursor (EC 3.4.24.36) (Cell surface protease)
 DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface
 DE endopeptidase).
 DE GP63-C1.
 OS Leishmania mexicana.
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 OX NCBI_TaxID=5665;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN=MNYC/BZ/62/M379;
 RX MEDLINE=93149206; PubMed=8426614;
 RA Medina-Acosta E., Kress R.E., Russell D.G.;
 RT "Structurally distinct genes for the surface protease of Leishmania
 RT mexicana are developmentally regulated";
 RL Mol. Biochem. Parasitol. 57:31-46(1993).
 CC -!- FUNCTION: Has an integral role during the infection of macrophages
 CC in the mammalian host.
 CC -!- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and
 CC P1', and basic residues at P2 and P3'. A model nonapeptide is
 CC cleaved at -Ala-Tyr-|-Leu-Lys-Lys-.
 CC -!- COPACTOR: Binds 1 zinc ion per subunit (By similarity).
 CC -!- DEVELOPMENTAL STAGE: Expressed in both the promastigote and the


```

RT "Empirical analysis of transcriptional activity in the Arabidopsis
RL genome.";
RL Science 302:842-846(2003).
CC -!- FUNCTION: Essential for proper development of leaves and floral
CC organs, and formation of axillary meristems.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=1;
CC Comment=A number of isoforms are produced. According to EST
CC sequences;
CC Name=1;
CC IsoId=O04379-1; Sequence=Displayed;
CC -!- TISSUE SPECIFICITY: Widely expressed at low levels.
CC -!- DEVELOPMENTAL STAGE: Expressed throughout all developmental
CC stages.
CC -!- SIMILARITY: Belongs to the argonaute family.
CC -!- SIMILARITY: Contains 1 PAZ domain.
CC -!- SIMILARITY: Contains 1 Piwi domain.
CC -!- CAUTION: Ref.2 (AAF79718) sequence differs from that shown due to
CC erroneous gene model prediction.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to licenses@isb-sib.ch).
CC -----
CC EMBL; U91995; AAC18440.1;
CC EMBL; AC007932; AAD49755.1;
CC EMBL; AC020889; AAF79718.1; ALT_SEQ.
CC EMBL; BT000941; AAN41341.1;
CC InterPro; IPR003100; PAZ.
CC InterPro; IPR003165; Piwi.
CC Pfam; PF02170; PAZ; 1.
CC Pfam; PF02171; Piwi; 1.
CC PROSITE; PS50821; PAZ; 1.
CC PROSITE; PS50822; PIWI; 1.
CC Developmental protein; Alternative splicing.
FT DOMAIN 391 501
FT DOMAIN 676 997
FT DOMAIN 13 104 GLY-RICH.
SQ SEQUENCE 1048 AA; 116190 MW; 3E5146343A09C541 CRC64;

Query Match 11.0%; Score 85.5; DB 1; Length 1048;
Best Local Similarity 25.7%; Pred.No.9.2;
Matches 36; Conservative 17; Mismatches 60; Indels 27; Gaps 3;

QY 27 GGGGNHNGGSSGPDSTLSI-----YQGSANAALALQSDARKSETTITQSGYG 76
Db 95 GYGGRGGGSPSPQRCQSPFELHQAQTSPTQAVSSQPTLSEVSPQVPEFTVLAQGF- 153
QY 77 NGADYVQQGADNSTIE-----LTQNGFRFNATIDOWNAKNSDITVQGYDQ 120
Db 154 EQLSVEQAGPSQAIOPISSKAFKFPMPRGKQSGKRCIVKANHFFAELPDKDLHHYDV 213
QY 121 LVTRVVTHEMAHASVMVRCV 140
Db 214 TITPEVTSRGVNRVAVNKQLV 233

```

Search completed: August 2, 2004, 14:49:28
Job time : 6.3 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:39:17 ; Search time 29.7 seconds

(without alignments)
1604.150 Million cell updates/sec

Title: US-09-543-407-14

Perfect score: 775
Sequence: 1 MKLLKVAFAAIVVSGSALA.....HASVMVRQVFGNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25.*

1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_virus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	684	88.3	152	Q33802	Q33802 salmonella
2	578.5	74.6	150	Q7X243	Q7X243 citrobacter
3	533	68.8	149	Q7X240	Q7X240 citrobacter
4	503.5	65.0	152	Q8CW63	Q8CW63 escherichia
5	438.5	56.6	150	Q7X237	Q7X237 enterobacte
6	385	49.7	76	Q54069	Q54069 salmonella
7	122	15.7	29	Q9S3J5	Q9S3J5 escherichia
8	109	14.1	179	Q33801	Q33801 salmonella
9	107.5	13.9	151	Q7X238	Q7X238 enterobacte
10	105	13.5	139	Q8EIH3	Q8EIH3 shewanella
11	104.5	13.5	152	Q7X241	Q7X241 citrobacter
12	104.5	13.5	160	Q8CW64	Q8CW64 escherichia
13	104	13.4	502	Q8EIH4	Q8EIH4 shewanella
14	103.5	13.4	151	Q7X244	Q7X244 citrobacter
15	103.5	13.4	151	Q7UC21	Q7UC21 shigella fl
16	103.5	13.4	160	Q83RU7	Q83RU7 shigella fl

17	101	13.0	362	16	Q8EV84	Q8EV84 mycoplasma
18	99.5	12.8	348	13	Q93397	Q93397 cyprinus ca
19	96.5	12.5	552	16	P96840	P96840 mycobacteri
20	96.5	12.5	552	16	Q7TW76	Q7TW76 mycobacteri
21	96.5	12.5	423	16	Q8VIY0	Q8VIY0 mycobacteri
22	96.5	12.5	1422	16	Q8EFU3	Q8EFU3 shewanella
23	96.5	12.5	3501	16	Q8YI06	Q8YI06 raistonia s
24	96.5	12.5	3552	16	Q8XSD6	Q8XSD6 raistonia s
25	96	12.4	154	16	Q89JIS	Q89JIS bradyrhizob
26	96	12.4	157	16	Q8EHG0	Q8EHG0 pseudomonas
27	96	12.4	606	3	Q9P3I9	Q9P3I9 neurospora
28	95.5	12.3	824	3	Q8NIV1	Q8NIV1 neurospora
29	95.5	12.3	1785	16	Q7V8S5	Q7V8S5 prochlorococ
30	95	12.3	1748	5	Q94821	Q94821 tetrahymena
31	94.5	12.2	191	3	Q8TFA6	Q8TFA6 saccharomyc
32	94	12.1	171	16	Q89JIS	Q89JIS bradyrhizob
33	94	12.1	368	16	Q8EWD6	Q8EWD6 mycoplasma
34	93	12.0	480	16	Q89EV2	Q89EV2 bradyrhizob
35	93	12.0	2174	16	Q92UU8	Q92UU8 rhizobium m
36	92	11.9	453	5	Q9N6M8	Q9N6M8 drosophila
37	92	11.7	301	10	Q9XIL0	Q9XIL0 arabidopsis
38	91	11.7	453	5	Q9NGF6	Q9NGF6 drosophila
39	91	11.7	453	5	Q9NGF7	Q9NGF7 drosophila
40	91	11.7	1222	5	Q9W4F0	Q9W4F0 drosophila
41	91	11.7	1222	5	Q8T8L9	Q8T8L9 drosophila
42	91	11.7	1351	5	Q8SX56	Q8SX56 drosophila
43	91	11.7	1354	5	Q8MPN4	Q8MPN4 drosophila
44	90.5	11.7	145	16	Q8U6N9	Q8U6N9 agrobacteri
45	90.5	11.7	191	10	Q7XDR3	Q7XDR3 oryza sativ

ALIGNMENTS

RESULT 1

O33802 : PRELIMINARY; PRT; 152 AA.
ID O33802 :
AC O33802 :
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE AgfA protein (Fragment).
DE AGFA.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OC NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98053981; PubMed=9393832;
RA Sukupolvi S.S., Lorentz R.G., Gordon J.I., Bian Z., Pfeifer J.D.,
RA Normark S.J., Rhen M.
RT "Expression of thin, aggregative fimbriae promotes interaction of
RT Salmonella typhimurium SR-11 with mouse small intestinal epithelial
RT cells.";
RL Infect. Immun. 65:5320-5325(1997).
DR EMBL; AJ000514; CAA04151.1; -;
FT NON_TER 152
FT SEQUENCE 152 AA; 15401 MW; 9DA7DADC2364B006 CRC64;
Query Match 88.3%; Score 684; DB 2; Length 152;
Best Local Similarity 89.4%; Pred. NO. 1.2e-47;
Matches 135; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

QY	1	MKLLKVAFAAIVVSGSALA	VVVGQGGNGNNGGSSGPDSTLSIYVGSANAALAQ	60
Db	1	MKLLKVAFAAIVVSGSALA	VVVGQGGNGNNGGSSGPDSTLSIYVGSANAALAQ	60
QY	61	SDARKSETTITGSGYNGADVGQ	GDADNSTIELTQNGFRNNATIDQWNAKNSDITVGYDQ	120
Db	61	SDARKSETTITGSGYNGADVGQ	GDADNSTIELTQNGFRNNATIDQWNAKNSDITVGYDQ	120
QY	121	LIVTRVTHEMAHNVVRQVFG	GNNTANQY 151	


```

RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RT Enterobacteriaceae isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158(2003).
DR EMBL; AJ515702; CAD56678.1; -.
SQ SEQUENCE 150 AA; 15112 MW; 508BB2D872DF15F3 CRC64;

Query Match 56.6%; Score 438.5; DB 2; Length 150;
Best Local Similarity 58.9%; Pred. No. 7e-28;
Matches 89; Conservative 28; Mismatches 33; Indels 1; Gaps 1;

QY 1 MKLLKVAFAAIVVSGSALAGVVPWGCGGNGHNGSSGPDSTLSIYQGSANAALALQ 60
DB 1 MKFIKVAALAIIVVSGSAMAGNIQ-GGWGHGHGGYGGNSILNIYQNGGNSALALQ 59

QY 61 SDARKSETTITQSGYNGADVQGGADNSTIELTQNGFNNATIDQWNAKNSDITVQGYDQ 120
DB 60 TDARNSVLNISQTGNGGADVQGGSDSSINLTQNGFGNSATLDQWNSKDSVMNVSYQGG 119

QY 121 LVTRVVTHEMAHASVMVRQVGFNNATANQY 151
DB 120 LNALVDQTASNSTVNTQIGFGNHATAHQY 150

RESULT 6
Q54069 PRELIMINARY; PRT; 76 AA.
AC Q54069;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE SEF17 fimbria (Fragment).
GN AGFA.
OS Salmonella enteritidis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=592;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SE30.
RA Cox J.M., Eglezos S., Woolcock J.B.;
RT "Virulence of Salmonella enteritidis in chickens correlates with
RT colony morphology and expression of SEF17 fimbriae.";
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
FR EMBL; U53207; AAA98671.1; -.
FT NON TER 1
FT NON TER 76
SQ SEQUENCE 76 AA; 7704 MW; 2FD5411241A7BCB1 CRC64;

Query Match 49.7%; Score 385; DB 2; Length 76;
Best Local Similarity 98.7%; Pred. No. 6.3e-24;
Matches 75; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 30 GNHNGGSSGPDSTLSIYQGSANAALALQSDARKSETTITQSGYNGADVQGGADNST 89
DB 1 GNHNGGSSGPDSTLSIYQGSANAALALQSDARKSETTITQSGYNGADVQGGADNST 60

QY 90 IELTQNGFRNNATIDQ 105
DB 61 IELTQNGFRNNATIDQ 76

RESULT 7
Q9S3J5 PRELIMINARY; PRT; 29 AA.
AC Q9S3J5;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 16, Last annotation update)
DE Curlin subunit monomer (Fragment).
GN CSGA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.

```

```

OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC TRANSPOSON=Insertion sequence IS1;
RX MEDLINE=99314153; PubMed=10386375;
RA La Regione R.M., Collighan R.J., Woodward M.J.;
RT "Non-cultivation of Escherichia coli O78:H80 isolates associated with
RT ISI insertions in csge and reduced persistence in poultry infection.";
RL FEMS Microbiol. Lett. 175:247-253(1999).
DR EMBL; AJ131756; CAB45380.1; -.
FT NON TER 29
FT NON TER 29
SQ SEQUENCE 29 AA; 2789 MW; E29D0FC07ABBE243 CRC64;

Query Match 15.7%; Score 122; DB 2; Length 29;
Best Local Similarity 89.7%; Pred. No. 0.003;
Matches 26; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPWGCGG 29
DB 1 MKLLKVAATAAIVVSGSALAGVVPQYGGG 29

RESULT 8
Q33801 PRELIMINARY; PRT; 179 AA.
AC Q33801;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE AgfB protein.
GN AGFB.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=98053981; PubMed=9393832;
RA Sukupolvi S.S., Lorentz R.G., Gordon J.I., Bian Z., Pfeifer J.D.,
RA Normark S.J., Rhen M.;
RT "Expression of thin, aggregative fimbriae promotes interaction of
RT Salmonella typhimurium SR-11 with mouse small intestinal epithelial
RT cells.";
RL Infect. Immun. 65:5320-5325(1997).
FR EMBL; AJ000514; CAA04150.1; -.
FT NON TER 1
FT NON TER 179
SQ SEQUENCE 179 AA; 19318 MW; A2BCCB648B3C0B0B CRC64;

Query Match 14.1%; Score 109; DB 2; Length 179;
Best Local Similarity 26.5%; Pred. No. 0.29;
Matches 36; Conservative 20; Mismatches 48; Indels 32; Gaps 4;

QY 16 GSALAGVVPWGCGGNGHNGSSGPDSTLSIYQGSANAALALQSDARKSETTITQSGY 75
DB 63 GSKLLSVISQ-ERGGNNRAKVDQAGNYFAYIEQTGNAN-----DASISQAY 109

QY 76 GNGADVQGGADNSTIELTQNGFNNATIDQWNAKNSDITV-----GQYQOLVT 123
DB 110 GNSAAI-----SAAIIQKSGNKANITQYGTQKTAVVVQKSHMAIQANITQYGTQKT 162

QY 124 RVVTHEVAHASVMVRQ 139
DB 163 AVVVQKSHMAIRVTQ 178

RESULT 9
Q7X238 PRELIMINARY; PRT; 151 AA.
AC Q7X238;
DT 01-OCT-2003 (TREMBLrel. 25, Created)
DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Nucleation component of curlin monomers.
GN CSGB.

```

```

OS Enterobacter sakazakii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Enterobacter.
OX NCBI_TaxID=28141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rec39;
RA Zogaj X., Bokranz W., Nimitz M., Romling U.;
RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RT Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158(2003).
DR EMBL; AJ515702; CAD56677.1; -.
SQ SEQUENCE 151 AA; 15985 MW; F0B82BD2A27882B7 CRC64;

Query Match 13.9%; Score 107.5; DB 2; Length 151;
Best Local Similarity 30.8%; Pred. No. 0.31;
Matches 36; Conservative 14; Mismatches 42; Indels 25; Gaps 4;

QY 10 AATVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQGSANAALQSDARKSETT 69
DQ 58 AQIRQSGSKLLSVSQ-----DGAGNRARD-----QSGTYNIAMIDQS-GNGNDAG 103

QY 70 ITQSGYNGADVQGGADNSTLTQGFNNATIDQWNAKNSDITVGYDQLVTRV 126
DQ 104 ITQDGYGNSA-----KIIQKSGNRRNITQYGTQKTAVVVQSQQWAIHV 149

RESULT 10
Q8EIH3 PRELIMINARY; PRT; 139 AA.
ID Q8EIH3
AC Q8EIH3;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Minor curlin subunit CsgB, putative.
GN S00866.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.
OX NCBI_TaxID=70863;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MR-1;
RX MEDLINE=22257686; PubMed=12369813;
RA Heidelberg J.F., Paulsen I.T., Nelson K.B., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Mathe B., Clayton R.A.,
RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Unayam L.A., White O., Wolf A.M.,
RA Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,
RA Mueller J., Khouri H., Gill J., Utterback T.R., McDonald L.A.,
RA Feldblyum T.V., Smith H.O., Venter J.C., Nelson K.H., Fraser C.M.;
RA "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis.";
RL Nat. Biotechnol. 20:1118-1123(2002).
DR EMBL; A2015532; AAN53942.1; -.
DR TIGR; S00866; -.
KW Complete proteome.
SQ SEQUENCE 139 AA; 14811 MW; 41EC1CFA76957920 CRC64;

Query Match 13.5%; Score 105; DB 16; Length 139;
Best Local Similarity 28.3%; Pred. No. 0.45;
Matches 32; Conservative 23; Mismatches 44; Indels 14; Gaps 4;

QY 39 SGPDSLTLSIYQGSANAALQSDARKSETTITQSGYNGADVQGGADNSTLTQNGFR 98
DQ 41 SGRENIDILVQGGTANQGVFQSGSDNS--AYVTQAGNDNISLVTQGTNNVQLLQVGAQ 99

QY 99 NNATIDQWNAKNSDITVGYDQLVTRVTHMAHASVMVRQGVGNNATANQY 151
DQ 100 NKASITQ-----IGN-DNLVQ---LNLQSGNFSIQIADGAAISITQY 139

```

```

RESULT 11
Q7X241 PRELIMINARY; PRT; 152 AA.
ID Q7X241
AC Q7X241;
DT 01-OCT-2003 (T-EMBLrel. 25, Created)
DT 01-OCT-2003 (T-EMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Nucleation component of curlin monomers.
GN C5GB.
OS Citrobacter freundii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Citrobacter.
OX NCBI_TaxID=546;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Pec4;
RA Zogaj X., Bokranz W., Nimitz M., Romling U.;
RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RT Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158(2003).
DR EMBL; AJ515701; CAD56674.1; -.
SQ SEQUENCE 152 AA; 16149 MW; D063A527D45D4329 CRC64;

Query Match 13.5%; Score 104.5; DB 2; Length 152;
Best Local Similarity 29.1%; Pred. No. 0.55;
Matches 32; Conservative 13; Mismatches 40; Indels 25; Gaps 3;

QY 16 GSALAGVVPQWGGGNGHNGSGSPDSTLSIYQGSANAALQSDARKSETTITQSGY 75
DQ 65 GSXLLSVISQ-----EGSGNRAKTDQT-----GSYNFAY-----IDQTGS 99

QY 76 GNGADVQGGADNSTLTQGFNNATIDQWNAKNSDITVGYDQLVTRV 125
DQ 100 SNDASIKQSGYNGTAVIIOKSGNKNANITQYGTQKTAVVVQSQQWAIHV 149

RESULT 12
Q8CW64 PRELIMINARY; PRT; 160 AA.
ID Q8CW64
AC Q8CW64;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Minor curlin subunit precursor.
GN C5GB OR C1305.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
DR EMBL; AE016759; AAN79778.1; -.
KW Complete proteome.
SQ SEQUENCE 160 AA; 16963 MW; 49F68448D979B986 CRC64;

Query Match 13.5%; Score 104.5; DB 16; Length 160;
Best Local Similarity 29.2%; Pred. No. 0.58;
Matches 35; Conservative 19; Mismatches 59; Indels 7; Gaps 4;

QY 9 FAATVVGSGAL--AGVVQWGGGNGHNGSGSPDSTLSIY-QYGSANAALQSDARK 65
DQ 42 FAVNELSKSFFNOAAII--GQAGTNNNAQLRQGGSKLLTVVAQEGSSNRA-KIDQTGY 97

QY 66 SETTITQSGYNGADVQGGADNSTLTQGFNNATIDQWNAKNSDITVGYDQLVTRV 125

```

Search completed: August 2, 2004, 14:54:37
Job time : 30.7 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:35:42 ; Search time 44.9 Seconds

(without alignments)
950.215 Million cell updates/sec

Title: US-09-543-407-16

Perfect score: 757

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVZQVGFNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	757	100.0	151	3	AAB36348
2	658	86.9	151	2	AAR74625
3	658	86.9	151	3	AAB36341
4	653	85.3	151	2	AAW23570
5	618	81.6	151	3	AAB36349
6	580	76.6	151	3	AAB36346
7	578	76.4	151	3	AAB36347
8	577	76.2	151	3	AAB36353
9	575	76.0	151	3	AAB36352
10	574	75.8	151	3	AAB36350
11	568	75.0	151	3	AAB36354
12	567	74.9	151	3	AAB36351
13	566	74.8	151	3	AAB36355
14	487	64.3	120	2	AAR62761
15	487	64.3	120	2	AAW23569
16	484	63.9	151	3	AAB36343
17	479	63.3	151	7	ABR82651
18	406	53.6	142	2	AAR52664
19	345	45.6	122	2	AAR52663
20	237	31.3	45	3	AAB36316
21	123	16.2	23	3	AAB36321
22	123	16.2	23	3	AAB36326
23	123	16.2	23	3	AAB36338
24	118.5	15.7	151	3	AAB36344
25	115	15.2	22	3	AAB36325

26	115	15.2	22	3	AAB36339	Aab36339	Salmonell
27	115	15.2	22	3	AAB36320	Aab36320	Salmonell
28	111.5	14.7	151	3	AAB36342	Aab36342	Salmonell
29	111	14.7	22	3	AAB36322	Aab36322	Salmonell
30	111	14.7	22	3	AAB36327	Aab36327	Salmonell
31	111	14.7	22	3	AAB36337	Aab36337	Salmonell
32	109	14.4	23	3	AAB36340	Aab36340	Salmonell
33	109	14.4	23	3	AAB36324	Aab36324	Salmonell
34	109	14.4	23	3	AAB36319	Aab36319	Salmonell
35	102	13.5	26	7	ABR82649	ABR82649	E. coli V
36	98	12.9	26	7	ABR82645	ABR82645	E. coli c
37	96.5	12.7	502	2	AAW32312	AAW32312	Leishman
38	96.5	12.7	2016	6	AAB36891	AAB36891	Plectre
39	96	12.7	19	3	AAB36323	Aab36323	Salmonell
40	96	12.7	19	3	AAB36336	Aab36336	Salmonell
41	96	12.7	19	3	AAB36328	Aab36328	Salmonell
42	95	12.5	24	7	ABR82647	ABR82647	E. coli c
43	95	12.5	764	6	AAB36890	AAB36890	Plectre
44	93	12.3	290	6	ABU44579	ABU44579	Protein e
45	92	12.2	23	3	AAB36331	Aab36331	Escherich

ALIGNMENTS

RESULT 1
AAB36348
ID AAB36348 standard; protein; 151 AA.
XX
AC AAB36348;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa::PT3#3 amino acid sequence SEQ ID NO:16.
XX
XX Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO2000060102-A2.
XX
XX 12-OCT-2000.
XX
XX 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
PA (UYVI-) UNIV VICTORIA.
PI White AP, Doran JL, Collison SK, Kay WW;
XX
XX WPI; 2000-672631/65.
DR N-PSDB; AAC64624.
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence which encodes foreign epitope or antigen, expresses recombinant Agfa protein useful for eliciting immune response in animal.
PS Disclosure; Page 136; 139pp; English.
XX
XX The present invention describes a recombinant agfa gene (1) where a segment of the gene has been replaced by a segment of a foreign DNA sequence which encodes a foreign epitope or antigen. Also described are: (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended assembly system of strains of Salmonella, Escherichia coli and Enterobacteriaceae for the production of fimbriae comprising recombinant Agfa, CsgA and Agfa-homologue fimbrin subunits, respectively; (2) directing recombination of a recombinant gene into the chromosome of the homologous species; (3) directing recombination of a recombinant gene back into the chromosome of the homologous species, replacing the native

CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbriae protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

XX
 SQ Sequence 151 AA;

Query Match 100.0%; Score 757; DB 3; Length 151;
 Best Local Similarity 100.0%; Pred. No. 2.2e-67;
 Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVDQLVTRVVTHEMAHAGSPDSTLSIYQYGSANAALALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVDQLVTRVVTHEMAHAGSPDSTLSIYQYGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITVQYGG 120
 DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITVQYGG 120

QY 121 NNAALVNQASDSSVMVRQVGFNNATANQY 151
 DB 121 NNAALVNQASDSSVMVRQVGFNNATANQY 151

RESULT 2
 AAR74625
 ID AAR74625 standard; protein; 151 AA.
 AC AAR74625;
 XX
 XX 25-MAR-2003 (revised)
 DT 26-JUN-1995 (first entry)
 XX
 XX Agfa sequence.
 DE
 XX
 KW Salmonella; Agfa; vaccine.
 XX
 OS Salmonella.
 XX
 XX WO9425598-A2.
 XX
 XX 10-NOV-1994.
 XX
 XX 26-APR-1994; 94WO-IB000207.
 XX
 XX 26-APR-1993; 93US-00054452.
 XX
 XX (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
 PA (KING/) KING J.
 XX
 XX Kay WW, Collinson SK, Clouthier SC, Doran JL;
 XX WPI; 1994-358275/44.
 DR N-PSDB; AAQ87467.
 XX
 XX Eliciting an immune response to Salmonella - using attenuated Salmonella
 PT strains, vector constructs, or compsns. contg. fimbrial type proteins.
 PT
 XX Disclosure; Fig 7B; 95pp; English.
 PS
 XX The Salmonella Agfa protein and DNA are used in vaccine and genetic

CC immunization compositions, respectively, to elicit an immune response to
 CC Salmonella in animals (e.g. food producing animals) and humans. (Updated
 CC on 25-MAR-2003 to correct FN field.)

XX
 SQ Sequence 151 AA;

Query Match 86.9%; Score 658; DB 2; Length 151;
 Best Local Similarity 90.1%; Pred. No. 1.6e-57;
 Matches 136; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVDQLVTRVVTHEMAHAGSPDSTLSIYQYGSANAALALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVDQLVTRVVTHEMAHAGSPDSTLSIYQYGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITVQYGG 120
 DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITVQYGG 120

QY 121 NNAALVNQASDSSVMVRQVGFNNATANQY 151
 DB 121 NNAALVNQASDSSVMVRQVGFNNATANQY 151

RESULT 3
 AAB36341
 ID AAB36341 standard; protein; 151 AA.
 AC AAB36341;
 XX
 XX 26-FEB-2001 (first entry)
 DT
 XX
 XX Salmonella enteritidis Agfa amino acid sequence SEQ ID NO:5.
 DE
 XX
 KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 XX
 XX WO200060102-A2.
 XX
 XX 12-OCT-2000.
 XX
 XX 05-APR-2000; 2000WO-CA000356.
 PF
 XX
 XX 05-APR-1999; 99US-0127888P.
 PR
 XX
 XX (UYVI-) UNIV VICTORIA.
 PA
 XX
 XX White AP, Doran JL, Collinson SK, Kay WW;
 FI WPI; 2000-672631/65.
 XX
 XX N-PSDB; AAC64617.
 DR
 XX
 XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 PT
 XX
 XX Disclosure; Page 135; 139pp; English.

The present invention describes a recombinant agfa gene (1) where a
 segment of the gene has been replaced by a segment of a foreign DNA
 sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or

CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fibrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fibrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

XX SQ Sequence 151 AA;

Query Match 86.9%; Score 658; DB 3; Length 151;
 Best Local Similarity 90.1%; Pred. No. 1.6e-57;
 Matches 136; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGYDQVTRVVTHEMAHSGDPDSTLSIYQGSANAALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGYVPOWGGGNGHNGSGSPDSTLSIYQGSANAALQ 60
 QY 61 SDARKSETTITQSGYNGADYVQGGADNSTIELTQNGFRNNATIDQWNAKNSDIIVGOYGG 120
 DB 61 SDARKSETTITQSGYNGADYVQGGADNSTIELTQNGFRNNATIDQWNAKNSDIIVGOYGG 120
 QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
 DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 4

AAW23570
 ID AAW23570 standard; protein; 151 AA.

XX AC AAW23570;
 XX DT 25-MAR-2003 (revised)
 XX DT 29-SEP-1997 (first entry)
 XX DE Salmonella enteritidis 27655-3b agfa.
 XX KW Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody.
 XX OS Salmonella enteritidis.

XX FH Key Location/Qualifiers
 XX FT Misc-difference 123
 XX FT /note= "Encoded by GCC"

XX PN US5635617-A.
 XX PD 03-JUN-1997.
 XX PF 26-APR-1994; 94US-00233788.
 XX PR 26-APR-1993; 93US-00054452.
 XX PA (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.

XX PI Collinson SK, Kay WM, Doran JL;
 XX DR WPI; 1997-309886/28.
 XX DR N-PSDB; AAT74142.

XX PT Isolated Salmonella gene agfa - used for diagnosis of Salmonella or
 XX PT enteropathogenic bacteria of the Enterobacteria family.

XX PS Example 2; Fig 7; 85pp; English.

XX CC The present sequence represents agfa encoded by the full agfa gene
 XX CC derived from Salmonella enteritidis 27655-3b. The nucleic acid can be

CC used to provide diagnostic assays for Salmonella and/or enteropathogenic
 CC bacteria of the family Enterobacteria. It can also be used to provide
 CC proteins and antibodies which can be used for assays. The nucleic acid
 CC sequence can be used to provide probes or primers which can specifically
 CC hybridise to nucleic acid molecules from greater than 99% of Salmonella
 CC strains that are pathogenic to warm-blooded animals relative to nucleic
 CC acid molecules from virtually all other microbial organisms. (Updated on
 CC 25-MAR-2003 to correct PF field.)

XX SQ Sequence 151 AA;

Query Match 86.3%; Score 653; DB 2; Length 151;
 Best Local Similarity 89.4%; Pred. No. 5.1e-57;
 Matches 135; Conservative 3; Mismatches 13; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGYDQVTRVVTHEMAHSGDPDSTLSIYQGSANAALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGYVPOWGGGNGHNGSGSPDSTLSIYQGSANAALQ 60
 QY 61 SDARKSETTITQSGYNGADYVQGGADNSTIELTQNGFRNNATIDQWNAKNSDIIVGOYGG 120
 DB 61 SDARKSETTITQSGYNGADYVQGGADNSTIELTQNGFRNNATIDQWNAKNSDIIVGOYGG 120
 QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
 DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 5

AAE36349
 ID AAB36349 standard; protein; 151 AA.

XX AC AAB36349;

XX DT 26-FEB-2001 (first entry)

XX DE Agfa::PT3#4 amino acid sequence SEQ ID NO:18.

XX KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
 XX KW vaccine; immune response; immunogen.

XX OS Salmonella enteritidis.

XX OS Escherichia coli.

XX OS Synthetic.

XX PN WO2000060102-A2.

XX PD 12-OCT-2000.

XX PF 05-APR-2000; 2000WO-CA000356.

XX PR 05-APR-1999; 99US-0127888P.

XX PA (UYVI-) UNIV VICTORIA.

XX PI White AP, Doran JL, Collinson SK, Kay WM;

XX DR WPI; 2000-672631/65.

XX DR N-PSDB; AAC64625.

XX PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 XX PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 XX PT protein useful for eliciting immune response in animal.

XX PS Disclosure; Page 136; 139pp; English.

XX CC The present invention describes a recombinant agfa gene (I) where a
 XX CC segment of the gene has been replaced by a segment of a foreign DNA
 XX CC sequence which encodes a foreign epitope or antigen. Also described are:
 XX CC (1) use of thin aggregative fimbriae (SEF17/7AF) nucleation depended
 XX CC assembly system of strains of Salmonella, Escherichia coli and
 XX CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 XX CC Agfa, CsgA and Agfa-homologue fimbrin subunits, respectively; (2)

CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX Sequence 151 AA;

Query Match 81.6%; Score 618; DB 3; Length 151;
 Best Local Similarity 76.6%; Pred. No. 1.6e-53;
 Matches 131; Conservative 0; Mismatches 0; Indels 40; Gaps 2;
 QY 1 MKLLKVAAPAAIVVSGSALAGV-----YDQLVTRVVTHEMAHAGS 40
 DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGSGSDPDYDQLVTRVVTHEMAH--- 57
 QY 41 PSTLSIYQYGSANAALALQSDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNN 100
 DB 58 -----ALQSDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNN 100
 QY 101 ATIDQWNAKNSDITVQYGGNNAALVNQTSASSVVMVRQVGFNNATANQY 151
 DB 101 ATIDQWNAKNSDITVQYGGNNAALVNQTSASSVVMVRQVGFNNATANQY 151

RESULT 6
 AAB36346
 ID AAB36346 standard; protein; 151 AA.
 AC AAB36346;
 XX
 XX 26-FEB-2001 (first entry)
 XX
 XX Agfa::PT3#1 amino acid sequence SEQ ID NO:12.
 DE
 DE *Salmonella*; agfa; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 KW
 OS *Salmonella enteritidis*.
 OS *Escherichia coli*.
 OS Synthetic.
 XX
 XX WO2000060102-A2.
 PN
 XX
 XX 12-OCT-2000.
 PD
 XX
 XX 05-APR-2000; 2000WO-CA000356.
 PF
 XX
 XX 05-APR-1999; 99US-0127888P.
 PR
 XX
 XX (UYVI-) UNIV VICTORIA.
 PA
 XX
 XX White AP, Doran JL, Collison SK, Kay WW;
 FI
 XX
 XX WPI; 2000-672631/65.
 DR
 DR N-PSDB; AAC64622.
 XX
 XX
 XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PF
 FT

PT protein useful for eliciting immune response in animal.
 XX
 XX Disclosure; Page 135; 139pp; English.
 PS
 XX The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
 CC Agfa, CSa and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX Sequence 151 AA;

Query Match 76.6%; Score 580; DB 3; Length 151;
 Best Local Similarity 80.1%; Pred. No. 9.8e-50;
 Matches 121; Conservative 7; Mismatches 23; Indels 0; Gaps 0;
 QY 1 MKLLKVAAPAAIVVSGSALAGVYDQLVTRVVTHEMAHAGSDPSTLSIYQYGSANAALALQ 60
 DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGSGSDPSTLSIYQYGSANAALALQ 60
 QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGG 120
 DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGG 120
 QY 121 NNAALVNQTSASSVVMVRQVGFNNATANQY 151
 DB 121 NNAALVNQTSASSVVMVRQVGFNNATANQY 151

RESULT 7
 AAB36347
 ID AAB36347 standard; protein; 151 AA.
 AC AAB36347;
 XX
 XX 26-FEB-2001 (first entry)
 XX
 XX Agfa::PT3#2 amino acid sequence SEQ ID NO:14.
 DE
 DE *Salmonella*; agfa; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 KW
 OS *Salmonella enteritidis*.
 OS *Escherichia coli*.
 OS Synthetic.
 XX
 XX WO2000060102-A2.
 PN
 XX
 XX 12-OCT-2000.
 PD
 XX
 XX 05-APR-2000; 2000WO-CA000356.
 PF
 XX

```

PR 05-APR-1999; 99JUS-0127888P.
XX (UYVI-) UNIV VICTORIA.
XX
XX White AP, Doran JL, Collison SK, Kay WW;
XX
XX WPI; 2000-672631/65.
XX
XX N-PSDB; AAC64623.
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
XX
XX Disclosure; Page 136; 139pp; English.
XX
XX The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SfP17/TAF) nucleation depended are
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell) the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
XX SEQ Sequence 151 AA;
XX
XX Query Match 76.4%; Score 578; DB 3; Length 151;
XX Best Local Similarity 80.8%; Pred. No. 1.5e-49;
XX Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0
XX
XX QY 1 MKLLKVAFAAIVSGSALAGVYDQIVTRVVTHEMAHSGPDSLSTIYQYGSANALALQ 60
XX ||||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 1 MKLLKVAFAAIVSGSALAGVVPQWGGGNHGGNGSGPDSLSTIYQYGSANALALQ 60
XX
XX QY 61 SDARKSETTITQSGYNGADVQGGADNSTIELTQNGFNFNATIDQWNAKNSDITVQYGG 120
XX ||||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 61 SDARKSETTITQSGYNGADVQGGADNSTIELTQNGFNFNATIDQWNAKNSDITVQYQDQ 120
XX
XX QY 121 NNALVNVQTASDSSVMVQVGFNNATANQY 151
XX : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 121 LVTRVVVTHEMAHASYMVRQVGFNNATANQY 151
XX
XX AC AAB36353;
XX
XX AC AAB36353;
XX
XX DT 26-FEB-2001 (first entry)
XX
XX DE Agfa::PT3#9 amino acid sequence SEQ ID NO:26.
XX
XX Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen.

```


CC the exemplification of the present invention
 XX Sequence 151 AA;
 SQ

Query Match 75.8%; Score 574; DB 3; Length 151;
 Best Local Similarity 80.8%; Pred. NO. 3.9e-49;
 Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVYDQLVTRVVTHEMAHASGPDSTLSIYQYGSANAALALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVYDQLVTRVVTHEMAHASGPDSTLSIYQYGSANAALALQ 60
 QY 61 SDARKSETTITQSGYNGADYVQGGADNSTIELTQNGFRNNATIDOWNAKNSDITVGYGG 120
 DB 61 SDARKSETTITQSGYNGADYVQGGADNSTIELTQNGFRNNATIDOWNAKNSDITVGYGG 120
 QY 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151
 DB 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151

RESULT 11
 AAB36354
 ID AAB36354 standard; protein; 151 AA.
 XX
 AC AAB36354;
 DT 26-FEB-2001 (first entry)
 DE Agfa::PT3#9 amino acid sequence SEQ ID NO:28.
 XX
 KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 FN WO2000060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay WW;
 XX
 DR WPI; 2000-672631/65.
 DR N-PSDB; AAC64630.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 PS Disclosure; Page 138; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologous fimbrin subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or

CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX Sequence 151 AA;
 SQ

Query Match 75.0%; Score 568; DB 3; Length 151;
 Best Local Similarity 80.1%; Pred. NO. 1.5e-48;
 Matches 121; Conservative 4; Mismatches 26; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVYDQLVTRVVTHEMAHASGPDSTLSIYQYGSANAALALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVYDQLVTRVVTHEMAHASGPDSTLSIYQYGSANAALALQ 60
 QY 61 SDARKSETTITQSGYNGADYVQGGADNSTIELTQNGFRNNATIDOWNAKNSDITVGYGG 120
 DB 61 SDARKSETTITQSGYNGADYVQGGADNSTIELTQNGFRNNATIDOWNAKNSDITVGYGG 120
 QY 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151
 DB 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151

RESULT 12
 AAB36354
 ID AAB36354 standard; protein; 151 AA.
 XX
 AC AAB36354;
 DT 26-FEB-2001 (first entry)
 DE Agfa::PT3#6 amino acid sequence SEQ ID NO:22.
 XX
 KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 FN WO2000060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay WW;
 XX
 DR WPI; 2000-672631/65.
 DR N-PSDB; AAC64627.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 PS Disclosure; Page 137; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:

CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant AgfA
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX Sequence 151 AA;
 SQ

Query Match 74.9%; Score 567; DB 3; Length 151;
 Best Local Similarity 80.8%; Pred. No. 1.9e-48;
 Matches 122; Conservative 5; Mismatches 24; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIIVVSGSALAGYVDQLVTRVTHMAHAGSPDSTLSIYQGSANAALALQ 60
 DB 1 MKLLKVAAPAAIIVVSGSALAGYVDQLVTRVTHMAHAGSPDSTLSIYQGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNATIDQWNAKNSDITVQYGG 120
 DB 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNATIDQWNAKNSDITVQYGG 120

QY 121 NNAALVNQASDSSVMVRQVGFNNATANQY 151
 DB 121 NNAALVNQASDSSVMVRQVGFNNATANQY 151

RESULT 13
 AAB36355
 ID AAB36355 standard; protein; 151 AA.
 XX
 AC AAB36355;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE AgfA::P73#10 amino acid sequence SEQ ID NO:30.
 XX
 KW Salmonella; agfA; chromosomal gene replacement; fimbrial; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 XX WO200060102-A2.
 XX
 XX 12-OCT-2000.
 XX
 XX 05-APR-2000; 2000WO-CA000355.
 XX
 XX 05-APR-1999; 99US-012788EP.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay WW;
 XX WPI; 2000-672631/65.
 DR

DR N-PSDB; AAC64631.
 XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant AgfA
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 139; 139pp; English.
 XX
 CC The present invention describes a recombinant agfA gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant AgfA
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX Sequence 151 AA;
 SQ

Query Match 74.8%; Score 566; DB 3; Length 151;
 Best Local Similarity 80.8%; Pred. No. 2.4e-48;
 Matches 122; Conservative 5; Mismatches 24; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIIVVSGSALAGYVDQLVTRVTHMAHAGSPDSTLSIYQGSANAALALQ 60
 DB 1 MKLLKVAAPAAIIVVSGSALAGYVDQLVTRVTHMAHAGSPDSTLSIYQGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNATIDQWNAKNSDITVQYGG 120
 DB 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNATIDQWNAKNSDITVQYGG 120

QY 121 NNAALVNQASDSSVMVRQVGFNNATANQY 151
 DB 121 NNAALVNQASDSSVMVRQVGFNNATANQY 151

RESULT 14
 AAR62761
 ID AAR62761 standard; protein; 120 AA.
 XX
 AC AAR62761;
 XX
 DT 27-AUG-2003 (revised)
 DT 25-MAR-2003 (revised)
 DT 26-JUN-1995 (first entry)
 XX
 XX AgfA sequence.
 XX
 KW Salmonella; AgfA; vaccine.
 XX
 OS Salmonella enteritidis.
 XX
 PN WO9425598-A2.
 XX
 XX 10-NOV-1994.
 PD

```

XX PF 26-APR-1994; 94WO-IB000207.
XX PR 26-APR-1993; 93US-00054452.
XX PA (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX PA (KING) KING J.
XX PI Kay WW, Collinson SK, Clouthier SC, Doran JL;
XX DR WPI; 1994-359275/44.
XX DR N-PSDB; AAQ73066.
XX PT Eliciting an immune response to Salmonella - using attenuated Salmonella
XX PT strains, vector constructs, or compens. contg. fimbrial type proteins.
XX FS Disclosure; Fig 7A; 95pp; English.
XX CC The sequence represents the Salmonella enteritis 27655-3b TnpHoA mutant
XX CC strain Agfa protein. The encoding DNA and isolated Agfa protein are used
XX CC in genetic immunization and vaccine compositions, respectively, to elicit
XX CC an immune response to Salmonella in animals (e.g. food producing animals)
XX CC and humans. (Updated on 25-MAR-2003 to correct FN field.) (Updated on 27-
XX CC AUG-2003 to correct OS field.)
XX SQ Sequence 120 AA;

Query Match 64.3%; Score 487; DB 2; Length 120;
Best Local Similarity 97.9%; Pred. No. 1.3e-40;
Matches 95; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 37 HASGPDSTLSIYQYGSANAALQSDARKSETTITQSGYNGADVGGQADNSTIELTQNG 96
Db :|||||
16 NSSGPDSTLSIYQYGSANAALQSDARKSETTITQSGYNGADVGGQADNSTIELTQNG 75

QY 97 FRNNATIDQWNAKNSDITVGYGGNNAALVNQTSADS 133
Db :|||||
76 FRNNATIDQWNAKNSDITVGYGGNNAALVNQTSADS 112

RESULT 15
AAW23569
ID AAW23569 standard; protein; 120 AA.
XX AC AAW23569;
XX DT 25-MAR-2003 (revised)
XX DT 29-SEP-1997 (first entry)
XX DE Salmonella enteritidis 27655-3b TnpHoA mutant agfa fragment.
XX KW Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody.
XX OS Salmonella enteritidis.
XX XX US5635617-A.
XX PN 03-JUN-1997.
XX PD 26-APR-1994; 94US-00233788.
XX PF 26-APR-1993; 93US-00054452.
XX PR (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX PA Collinson SK, Kay WW, Doran JL;
XX PI WPI; 1997-309886/28.
XX DR N-PSDB; AAT74141.
XX PT Isolated Salmonella gene agfa - used for diagnosis of Salmonella or
XX PT enteropathogenic bacteria of the Enterobacteria family.

```

```

PS Example 2; Fig 7; 85pp; English.
XX The present sequence represents an agfa fragment encoded by an agfa gene
XX CC fragment derived from Salmonella enteritidis 27655-3b TnpHoA mutant
XX CC strain. The nucleic acid can be used to provide diagnostic assays for
XX CC Salmonella and/or enteropathogenic bacteria of the family Enterobacteria.
XX CC It can also be used to provide proteins and antibodies which can be used
XX CC for assays. The nucleic acid sequence can be used to provide probes or
XX CC primers which can specifically hybridise to nucleic acid molecules from
XX CC greater than 99% of Salmonella strains that are pathogenic to warm-
XX CC blooded animals relative to nucleic acid molecules from virtually all
XX CC other microbial organisms. (Updated on 25-MAR-2003 to correct PF field.)
XX SQ Sequence 120 AA;

Query Match 64.3%; Score 487; DB 2; Length 120;
Best Local Similarity 97.9%; Pred. No. 1.3e-40;
Matches 95; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 37 HASGPDSTLSIYQYGSANAALQSDARKSETTITQSGYNGADVGGQADNSTIELTQNG 96
Db :|||||
16 NSSGPDSTLSIYQYGSANAALQSDARKSETTITQSGYNGADVGGQADNSTIELTQNG 75

QY 97 FRNNATIDQWNAKNSDITVGYGGNNAALVNQTSADS 133
Db :|||||
76 FRNNATIDQWNAKNSDITVGYGGNNAALVNQTSADS 112

Search completed: August 2, 2004, 14:48:25
Job time : 45.9 secs

```

This Page Blank (uspto)

Db 1 MLLKLVAFALVSGSALAGVVPQWGGGNGNGSGSPDSTLSIYQYGSANAALAQ 60
 QY 61 SPARKSETTITQSGYNGADVGQGDADNSTIETQNGFRNNATIDQWNAKNSDITVGOYGG 120
 Db 61 SPARKSETTITQSGYNGADVGQGDADNSTIETQNGFRNNATIDQWNAKNSDITVGOYGG 120
 QY 121 NNAALVNQASDSSVMVROVGFNNATANQY 151
 Db 121 NNPALVNQASDSSVMVROVGFNNATANQY 151

RESULT 2

US-08-233-788A-57
 ; Sequence 57, Application US/08233788A
 ; Patent No. 5635617
 ; GENERAL INFORMATION:
 ; APPLICANT: Doran, James L.
 ; APPLICANT: Kay, William W.
 ; APPLICANT: Collinson, Karen S.
 ; APPLICANT: Clouthier, Sharon C.
 ; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
 ; TITLE OF INVENTION: OF SALMONELLA
 ; NUMBER OF SEQUENCES: 61
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Seed and Berry
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue
 ; CITY: Seattle
 ; STATE: Washington
 ; COUNTRY: U.S.A.
 ; ZIP: 98104-7092
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/233,788A
 ; FILING DATE: 26-APR-1994
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: King, Joshua
 ; REGISTRATION NUMBER: 35,570
 ; REFERENCE/DOCKET NUMBER: 920043.403C2
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (206) 622-4900
 ; TELEFAX: (206) 682-6031
 ; TELEX: 3723836 SEEDANBERRY
 ; INFORMATION FOR SEQ ID NO: 57:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 120 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-233-788A-57

Query Match 64.3%; Score 487; DB 1; Length 120;
 Best Local Similarity 97.9%; Pred. No. 2.5e-43;
 Matches 95; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 37 HASGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGADVGQGDADNSTIETQNG 96
 Db 16 NSSGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGADVGQGDADNSTIETQNG 75
 QY FNNATIDQWNAKNSDITVGOYGGNNAALVNQASDS 133
 Db 76 FNNATIDQWNAKNSDITVGOYGGNNAALVNQASDS 112

RESULT 3

US-09-543-681A-5434
 ; Sequence 5434, Application US/09543681A
 ; Patent No. 6605709
 ; GENERAL INFORMATION:

; APPLICANT: GARY BRETON
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
 ; FILE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 2709.1002-001
 ; CURRENT APPLICATION NUMBER: US/09/543,681A
 ; CURRENT FILING DATE: 2000-04-05
 ; PRIOR APPLICATION NUMBER: US 60/128,706
 ; PRIOR FILING DATE: 1999-04-09
 ; NUMBER OF SEQ ID NOS: 8344
 ; SEQ ID NO 5434
 ; LENGTH: 2315
 ; TYPE: PRT
 ; ORGANISM: Proteus mirabilis
 ; US-09-543-681A-5434

Query Match 11.6%; Score 88; DB 4; Length 2315;
 Best Local Similarity 28.0%; Pred. No. 6;
 Matches 37; Conservative 15; Mismatches 54; Indels 26; Gaps 6;

QY 24 DQLVTRVWTHEMAHAGSPD-STLSIYQYGSANAALALQSDARKSETTITQSGYNGADYV 82
 Db 1700 DVTVTITLNDAGHAGAPDYSLAGSQH---DKIALHGDGK---VTYDGASTVTVI 1752
 QY 83 QGADNS-----TIELTQNGFRNNATIDQWNAKNSDITVGOYGG--NNAALVNQASDS 133
 Db 1753 PAGSKSVSFIVDPTMEANQAF-----NAEGMEKVVAITITGTSNNATAVTDIVNNA 1803
 QY 134 SVMVROVGFNN 145
 Db 1804 GASATGVYDGN 1815

RESULT 4

US-09-477-135A-131
 ; Sequence 131, Application US/09477135A
 ; Patent No. 6572865
 ; GENERAL INFORMATION:
 ; APPLICANT: Nano, Francis
 ; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
 ; TITLE OF INVENTION: immunostimulatory Peptides
 ; FILE REFERENCE: 52888
 ; CURRENT APPLICATION NUMBER: US/09/477,135A
 ; CURRENT FILING DATE: 2000-01-03
 ; PRIOR APPLICATION NUMBER: 08990823
 ; PRIOR FILING DATE: 1997-12-15
 ; PRIOR APPLICATION NUMBER: US 96/10375
 ; PRIOR FILING DATE: 1996-06-14
 ; PRIOR APPLICATION NUMBER: 60/000,254
 ; PRIOR FILING DATE: 1995-06-15
 ; NUMBER OF SEQ ID NOS: 169
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 131
 ; LENGTH: 943
 ; TYPE: PRT
 ; ORGANISM: Mycobacterium tuberculosis
 ; US-09-477-135A-131

Query Match 11.4%; Score 86.5; DB 4; Length 943;
 Best Local Similarity 24.8%; Pred. No. 2.4;
 Matches 39; Conservative 22; Mismatches 73; Indels 23; Gaps 6;

QY 7 AFAAAVWSG-----SALAGVYDQLVTRVWTHEMAHAGSPDSTLS-IYQYGSANAA 56
 Db 96 ALEFVSLVSNLLGQNPAPATAEAYEQMWADQVAMFGYHAGAAASALTFFQALPT 155
 QY 57 L-----ALQSDARKSETTITQSGYNGADVGQGDADNSTIETQNGFRNNATIDQWNAKNSD 112
 Db 156 VAGGGALVSAAAAAQVTTTRVFRNLGL-ANVGEGN-----VGNVGNVFNLCGANIGN 207
 QY 113 ITVGOYGGNNAALVNQASDSSVMVROVGFNNATAN 149
 Db 208 IGSNGTSSNGISGFGN-VGPGLTAALNNIGFGNTGSNN 243

```
RESULT 5
US-09-252-991A-26438
; Sequence 26438, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 26438
; LENGTH: 1415
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-26438

Query Match 11.2%; Score 85; DB 4; Length 1415;
Best Local Similarity 25.6%; Pred. No. 6.1; Indels 52; Gaps 7;
Matches 43; Conservative 19; Mismatches 54;

QY 24 DQLVT---RVVTHEM-----ALASGPDSTLSIVQYGSANA-----ALA 58
DB 357 DQATGTWRLSHEVSVAAGCAAGGSGQVHVAQGPAGANASDNGVTIVQQPAVD 416
QY 59 LOSDARKSETTITQSCYG-----NGADVQGGADNSTIELTQN-----GFR 98
DB 417 LAAGANGTSVAVQSQSGANGISGVVQSQNGANIGAGASDISVVQSQNSPNIGSGVN 476
QY 99 NNATIDOWNAXN----SDIVGVQYGGNNALVNOTADSSVMVRQVG 141
DB 477 GVTVVQSQNGANIGSGASGITVQV--SQNGANIGSGASISVVQSQSG 522

RESULT 6
US-08-194-290-7
; Sequence 7, Application US/08194290
; Patent No. 5500353
; GENERAL INFORMATION:
; APPLICANT: Smit, John
; APPLICANT: Bingle, Wade H
; TITLE OF INVENTION: Bacterial surface protein expression
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shiesinger, Arkwright & Garvey
; STREET: 3000 South Eads Street
; CITY: Arlington
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/194,290
; FILING DATE: 09-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Garvey, George A
; REGISTRATION NUMBER: 17737
; REFERENCE/DOCKET NUMBER: 5946-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-684-5600
; TELEFAX: 703-836-5288
; INFORMATION FOR SEQ ID NO: 7:

Query Match 11.2%; Score 82; DB 1; Length 1026;
Best Local Similarity 24.8%; Pred. No. 7.9;
Matches 34; Conservative 28; Mismatches 55; Indels 20; Gaps 7;

QY 11 AIVVSGSALAGVVDQIVTRVTHEMAHA---SGPSTLSIYQYGSNAALALQSDARKSE 67
DB 463 AAATAGATVAGRVNGAVT--ITDSAAASATTAGKATVTLGSGAA---TIDSSAL--- 513
QY 68 TTITQSGYGNADVQGA-----DNSTIELTONGFERNATI-DQWNAKNSDITVQYGGN 121
DB 514 TTVNLSGTGTSIGRGALTATPTANTLTINVNGLTTTGATIDSEAAADGGTTINAGS 573
QY 122 NAA--LVNQTASDSSVM 136
DB 574 TASSTIASLVAADATTL 590

RESULT 7
US-08-614-377A-7
; Sequence 7, Application US/08614377A
; Patent No. 5976864
; GENERAL INFORMATION:
; APPLICANT: Smit, John
; APPLICANT: Bingle, Wade H.
; APPLICANT: No. 5976864ellini, John F.
; TITLE OF INVENTION: EXPRESSION AND SECRETION OF
; TITLE OF INVENTION: HETEROLOGOUS
; TITLE OF INVENTION: POLYPEPTIDES FROM CAULOBACTER
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson PC
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/614,377A
; FILING DATE: 12-MAR-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/194,290
; FILING DATE: 09-FEB-1994
; CLASSIFICATION: 435
; APPLICATION NUMBER: US 07/895,367
; FILING DATE: 09-JUNE-1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Tsao, Y. Rocky
; REGISTRATION NUMBER: 34053
; REFERENCE/DOCKET NUMBER: 08106/002001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-5070
; TELEFAX: 617-542-8906
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1026 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
```


; APPLICANT: Weltzin, Richard A.
; APPLICANT: Guy, Bruno
; TITLE OF INVENTION: LT and CT in Parenteral Immunization
; TITLE OF INVENTION: Methods Against Helicobacter Infection
; FILE REFERENCE: 06132/055002
; CURRENT APPLICATION NUMBER: US/09/336,115C
; CURRENT FILING DATE: 1999-06-18
; PRIOR APPLICATION NUMBER: US 09/100,258
; PRIOR FILING DATE: 1998-06-19
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 691
; TYPE: PRT
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)...(18)
US-09-336-115C-12

Query Match 10.7%; Score 81; DB 4; Length 691;
Best Local Similarity 19.0%; Pred. No. 5.7;
Matches 33; Conservative 27; Mismatches 66; Indels 48; Gaps 5;

QY 8 AFAIVSGSALAGVYDQVLTTRV-----VTHEMAHAGPD 42
DB 100 AQAVALNALAAVGMQVIALFIGCGPGFTNNQSYSGFNTPALNGTTTCNQAYGTGN 159
QY 43 STLSTIYQYGSANAALQSDARKSETTITQSYGNGADVQGGADNSITLTQNGFR--- 98
DB 160 GILSIDEYQKLNQAYQI-----TQALNQGGMPALNTTKGVNIOQTNYRTTQ 213
QY 99 NNATIDQWNAKNSDITVQYGG-----NNAA-LVNOTASDSSVMVRQ 139
DB 214 NMIIEHYTENGKEIPVSYSGSGSPFTIQLTYHNNAENLIQQAATIMQVLITQ 267

RESULT 12
US-09-336-115C-6
; Sequence 6, Application US/09336115C
; Patent No. 6576244
; GENERAL INFORMATION:
; APPLICANT: Weltzin, Richard A.
; APPLICANT: Guy, Bruno
; TITLE OF INVENTION: LT and CT in Parenteral Immunization
; TITLE OF INVENTION: Methods Against Helicobacter Infection
; FILE REFERENCE: 06132/055002
; CURRENT APPLICATION NUMBER: US/09/336,115C
; CURRENT FILING DATE: 1999-06-18
; PRIOR APPLICATION NUMBER: US 09/100,258
; PRIOR FILING DATE: 1998-06-19
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 745
; TYPE: PRT
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)...(20)
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 721
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-336-115C-6

Query Match 10.7%; Score 81; DB 4; Length 745;
Best Local Similarity 23.1%; Pred. No. 6.4;
Matches 27; Conservative 18; Mismatches 68; Indels 4; Gaps 2;

QY 31 VTHEMAHAGPSTLSIYQYGSANAALQSDARKSETTITQSYGNGADVQGGADNSTI 90
DB 152 ITCNRFSTGPGKGSIDEFKLNENAYQIIQOALKNQSGFPFLG-GNGTKV---SVNRY 207

QY 91 ELTQNGFRNNATIDQWNAKNSDITVQYGGNNAALVNQTSADSSVMVRQVGFNNAT 147
DB 208 ECRQTADINGVYQFCCKARKGSSSSNGGSGSTQTATTATQDGVITTTNNKAT 264

RESULT 13
US-09-206-942-69
; Sequence 69, Application US/09206942
; Patent No. 6432669
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High
; TITLE OF INVENTION: Molecular Weight Proteins
; FILE REFERENCE: 1038-861 MIS-7b
; CURRENT APPLICATION NUMBER: US/09/206,942
; CURRENT FILING DATE: 1998-12-08
; EARLIER APPLICATION NUMBER: 09/167,568
; EARLIER FILING DATE: 1998-10-07
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 69
; LENGTH: 1095
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-206-942-69

Query Match 10.6%; Score 80.5; DB 4; Length 1095;
Best Local Similarity 26.4%; Pred. No. 12;
Matches 39; Conservative 22; Mismatches 60; Indels 27; Gaps 7;

QY 15 SGSALAGVYDQ--LVTRVVTH---EMAHAGPSTLSIYQYGSANAALQSDARKSET 68
DB 727 TGSILGSISSGSGVTLTATEGALAVSNISG--NTVVTANSGLATTLAGSTIKGTESVT 784

QY 69 TITQSGYNGADVQGGADNSITEL-----TQNGFRNNATIDQWNAKNSDITVQYGGN 121
DB 785 TSSQSG-----DIGTISGTVKATESLTQNSKIKATTEANVTATGTIGTISG 839
QY 122 NAALVNQTSADSSVMVRQVGFNNATAN 149
DB 840 NTVVTANAGDLTV-----GNGAEIN 860

RESULT 14
US-08-038-682-2
; Sequence 2, Application US/08038682
; Patent No. 5549897
; GENERAL INFORMATION:
; APPLICANT: BARENKAMP, STEPHEN J
; APPLICANT: ST. GENE III, JOSEPH W
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/038,682
; FILING DATE: 16-MAR-1993
; CLASSIFICATION: 424

```

; ATTORNEY/AGENT INFORMATION:
; NAME: BERKSTRESSER, JERRY W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-293
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1536 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-038-682-2

Query Match 10.6%; Score 80.5; DB 1; Length 1536;
Best Local Similarity 26.4%; Pred. No. 20;
Matches 39; Conservative 22; Mismatches 60; Indels 27; Gaps 7;

QY 15 SGSALAGVYDQ--LVTRVYTH---EMAHASGPDSTLSIYQGSANAALALQS-DARKSET 68
Db 1168 TGSILGGIESSSGSVTLTATGALAVSNISG--NTVTVTANSGLTTLTLAGSTIKGTESVT 1225

QY 69 TITQSGYNGADYVGQADNSTIEL-----TONGFRNNATIDOWNAKNSDITVQYQYGN 121
Db 1226 TSSQSG-----DIGGTISGTVVEVKATESLTTSQNSKIKAATTCGEANVTSATGTIGGTISG 1280

QY 122 NAALVNQATASDSSVMVRQVFGNNATAN 149
Db 1281 NTVNVTANAGDLTV-----GNGAEIN 1301

Search completed: August 2, 2004, 14:58:32
Job time : 12 secs

```

RESULT 15

```

US-08-302-832-2
; Sequence 2, Application US/08302832
; Patent No. 5603938
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5603938-Typeable Haemophilus
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/302,832
; FILING DATE: 16-SEP-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US pct/us93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-404
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 2:

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:54:48 ; Search time 36.8 Seconds

(without alignments)
1287.123 Million cell updates/sec

Title: US-09-543-407-16

Perfect score: 757

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVQVFGNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubppaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubppaa/ECT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubppaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubppaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubppaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubppaa/ECTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubppaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubppaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubppaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubppaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubppaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubppaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubppaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubppaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubppaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubppaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubppaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubppaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	481	63.5	151	12	US-09-741-873B-4
2	481	63.5	151	12	US-09-741-873B-4
3	414	54.7	131	12	US-09-741-873B-2
4	113.5	15.0	445	15	US-10-369-493-20638
5	93	12.3	290	12	US-10-282-122A-72503
6	89.5	11.8	1862	12	US-10-282-122A-49757
7	89	11.8	182	14	US-10-238-075-1549
8	88.5	11.7	597	9	US-09-793-306-146
9	86.5	11.4	154	16	US-10-437-963-162284
10	86.5	11.4	943	9	US-09-996-634-131
11	86.5	11.4	943	10	US-09-997-182-131
12	86.5	11.4	943	10	US-09-997-181-131
13	86.5	11.4	3300	12	US-10-282-122A-64369
14	85	11.4	1554	16	US-10-383-930-38
15	85	11.2			

Sequence 107204, A
Sequence 49854, A
Sequence 1, Appli
Sequence 34, Appli
Sequence 1, Appli
Sequence 9411, Ap
Sequence 11, Appli
Sequence 10, Appli
Sequence 70017, A
Sequence 16, Appli
Sequence 33, Appli
Sequence 13173, A
Sequence 20096, A
Sequence 5343, Ap
Sequence 7, Appli
Sequence 109, App
Sequence 73678, A
Sequence 70586, A
Sequence 49773, A
Sequence 12, Appli
Sequence 22, Appli
Sequence 6, Appli
Sequence 5883, A
Sequence 73345, A
Sequence 54094, A
Sequence 65, Appli
Sequence 2, Appli
Sequence 63, Appli
Sequence 146426, A

US-10-437-963-107204
US-10-282-122A-49854
US-09-984-334-1
US-10-004-115A-34
US-10-327-108-1
US-10-156-761-9411
US-10-185-990-11
US-10-185-990-11
US-10-282-122A-70017
US-09-996-194-16
US-10-164-966-33
US-10-156-761-13173
US-10-369-493-20096
US-10-369-493-5343
US-09-379-931-7
US-10-223-597-7
US-09-769-787-109
US-10-282-122A-73678
US-10-282-122A-70586
US-10-282-122A-49773
US-08-834-666A-12
US-08-834-666A-22
US-08-834-666A-6
US-10-282-122A-5883
US-10-282-122A-73345
US-10-282-122A-54094
US-10-193-764-65
US-10-092-880-2
US-10-193-764-63
US-10-437-963-146426

ALIGNMENTS

RESULT 1

US-09-741-873B-4
; Sequence 4, Application US/09/41873B
; Publication No. US20020081722A1
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; CURRENT FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-4

Query Match 63.5%; Score 481; DB 12; Length 151;
Best Local Similarity 66.2%; Pred. No. 2.1e-44;
Matches 100; Conservative 19; Mismatches 32; Indels 0; Gaps 0;

Qy 1 MKLLKVAFAAIVVSGSALAGVYDQIVTRVVTWTHMAHSGPDSTLSIYQGSANAALQ 60

```

Db      1  MLLKVAALAAIFVSGSAGVVPVQVGGGNGHGGGNSGPNSELNIYQVGGNSALALQ 60
          :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      61  SPARKSETTITQSGYNGADVQGGADNSTIETQNGFRNATIDQWNAKNSDITVGGYGG 120
          :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db      61  TDARNSDLITTHGGGNGADVQGGSDSDIDLTQRGFGNSATLDQWNGKNSMTVKQFGG 120
          :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Qy      121  NNAALVNOTASDSSVMVRQVGFNNATANQY 151
          :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db      121  GNGAAVDQTASNSVNVTVQVGFNNATAHQY 151
          :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

RESULT 2

```

US-09-741-873B-4
; Sequence 4, Application US/09741873B
; Publication No. US20040096965A9
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; CURRENT FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-4

```

```

Query Match      63.5%; Score 481; DB 12; Length 151;
Best Local Similarity 66.2%; Pred. No. 2.1e-44;
Matches 100; Conservative 19; Mismatches 32; Indels 0; Gaps 0;

```

```

Qy      1  MLLKVAAPAAIVSGSALAGYVDLVTRVVTHEMAHASGDDSTLSIYQVGSANAALALQ 60
          :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db      1  MLLKVAALAAIFVSGSAGVVPVQVGGGNGHGGGNSGPNSELNIYQVGGNSALALQ 60
          :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Qy      61  SPARKSETTITQSGYNGADVQGGADNSTIETQNGFRNATIDQWNAKNSDITVGGYGG 120
          :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db      61  TDARNSDLITTHGGGNGADVQGGSDSDIDLTQRGFGNSATLDQWNGKNSMTVKQFGG 120
          :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Qy      121  NNAALVNOTASDSSVMVRQVGFNNATANQY 151
          :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db      121  GNGAAVDQTASNSVNVTVQVGFNNATAHQY 151
          :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

RESULT 3

```

US-09-741-873B-2
; Sequence 2, Application US/09741873B
; Publication No. US20020081722A1
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; CURRENT FILING DATE: 2003-04-04

```

```

; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-2

```

```

Query Match      54.7%; Score 414; DB 12; Length 131;
Best Local Similarity 69.9%; Pred. No. 3.6e-37;
Matches 79; Conservative 17; Mismatches 17; Indels 0; Gaps 0;

```

RESULT 4

```

US-09-741-873B-2
; Sequence 2, Application US/09741873B
; Publication No. US20040096965A9
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; CURRENT FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-2

```

```

Query Match      54.7%; Score 414; DB 12; Length 131;
Best Local Similarity 69.9%; Pred. No. 3.6e-37;
Matches 79; Conservative 17; Mismatches 17; Indels 0; Gaps 0;

```

```

Qy      39  SGPDSTLSIYQVGSANAALQSDARKSETTITQSGYNGADVQGGADNSTIETQNGFR 98
          :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db      19  SGENSELTVQYGGNSALALQTDARNSDLTTHGGGNGADVQGGSDSDIDLTQRGFG 78
          :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Qy      99  NNATIDQWNAKNSDITVQYGGNNAALVNQATSDSSVMVRQVGFNNATANQY 151
          :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db      79  NSATLDQWNGKNSMTVKQFGGNGAAVQDTASNSVNVTVQVGFNNATAHQY 131
          :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```



```
Db 19 SGPNSLTYQYGGNSALALQTDARNSDLTITQGGNGADVCGGSDSSIDLTORFG 78
QY 99 NNATIDOWNAKNSDITVCGYCGNAAALVNQTSASSVWVROVGFNNATANQY 151
Db 79 NSATLDOWNKXSEMTVKQFGGGAADVQDTASNSVNVTVQVGFNNATAHQY 131

RESULT 5
US-10-369-493-20638
; Sequence 20638, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 20638
; LENGTH: 445
; TYPE: PRT
; ORGANISM: Rhodospseudomonas palustris
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(445)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-20638

Query Match 15.0%; Score 113.5; DB 15; Length 445;
Best Local Similarity 25.0%; Pred. No. 0.0013; Indels 57; Gaps 5;
Matches 46; Conservative 18; Mismatches 63;

QY 3 LKLVAAFAAIVVSGSALAG---VYDQLVTRVVTTHMAHAGSPDSTLSIYQYGSANAALA 58
Db 5 LFFVTASVLVSSSAAFAADSNVY-----LNQTNDOQANITQSGNSVGA 52
QY 59 -----LQSDARKSETTI-----TQSGYNGADVCGGADNSTIELT 93
Db 53 FNGNSGFLQENGLTSGANLLTVKQSGNSVGRDIQKQSGAGNSAALFQEGTQSDVELQ 112
QY 94 QNGFERNATIDOWNAKN-----SDITVGYGNNAAALVNQTSASSVWV 137
Db 113 QTGTSNGAVPSGWNWINDPGVFNKITQDSSNSGSKSVIQQGKNVFSIKQNTGNSTSV 172
QY 138 RQVG 141
Db 173 NQIG 176

RESULT 6
US-10-282-122A-72503
; Sequence 72503, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.

Db 19 SGPNSLTYQYGGNSALALQTDARNSDLTITQGGNGADVCGGSDSSIDLTORFG 78
QY 99 NNATIDOWNAKNSDITVCGYCGNAAALVNQTSASSVWVROVGFNNATANQY 151
Db 79 NSATLDOWNKXSEMTVKQFGGGAADVQDTASNSVNVTVQVGFNNATAHQY 131

RESULT 7
US-10-282-122A-49757
; Sequence 49757, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.

QY 2 KLVAAFAAIVVSGSALAGVYDQLVTRVVTTHMAHAGSPDSTLSIYQYGSANAALA-- 59
Db 3 KLVK-----SIKSGSFFRAVLSTET-VRTAGEEHQSLSSTVALGRTLIANQLAANQ 56
QY 60 QSDARKSETTITQSGYNGADVCGGADNSTIELTQNGFRNNATIDOWNAKNSDITVGYG 119
Db 57 KGDGSKVTVKVIGDSSFGHILSVADTKGN-----VKGYIQNAGVDVVKKTASGEVIVGPFM 110
QY 120 GNNAAALVNQTSASSVWVROVGFNNATAN 149
Db 111 GNG-----QFVITDYGTFNPTSS 130

Query Match 12.3%; Score 93; DB 12; Length 290;
Best Local Similarity 26.7%; Pred. No. 0.12;
Matches 40; Conservative 23; Mismatches 63; Indels 24; Gaps 5;

QY 2 KLVAAFAAIVVSGSALAGVYDQLVTRVVTTHMAHAGSPDSTLSIYQYGSANAALA-- 59
Db 3 KLVK-----SIKSGSFFRAVLSTET-VRTAGEEHQSLSSTVALGRTLIANQLAANQ 56
QY 60 QSDARKSETTITQSGYNGADVCGGADNSTIELTQNGFRNNATIDOWNAKNSDITVGYG 119
Db 57 KGDGSKVTVKVIGDSSFGHILSVADTKGN-----VKGYIQNAGVDVVKKTASGEVIVGPFM 110
QY 120 GNNAAALVNQTSASSVWVROVGFNNATAN 149
Db 111 GNG-----QFVITDYGTFNPTSS 130

; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 72503
; LENGTH: 290
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-282-122A-72503
```

```

; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 49757
; LENGTH: 1862
; TYPE: PRT
; ORGANISM: Burkholderia fungorum
US-10-282-122A-49757

Query Match      11.8%; Score 89.5; DB 12; Length 1862;
Best Local Similarity 26.8%; Pred. No. 3.7;
Matches 51; Conservative 16; Mismatches 66; Indels 57; Gaps 8;

QY      7 ARAAAT-----VSSGALAGYDQVTRVTHMAHA--SGPDTLSIYQYGSANA 55
Db      1278 AASATATANGDVTVASAGKAAAGTTGVTGALNDINVAGATIESNGDAVLNA-QQGSLSNA 1336

QY      56 ALALQSDARKSETTITQSGYNGADVGGADNS-----TIELTQNGFRNNATIDOWN--- 107
Db      1337 TGGINSGA---ELTIT-----TGLDLSLGASTSAVGCDVTLNAGRNALINGTLVGQNGYI 1388

QY      108 AKNSDIT-----VGQYGNNAALVNQATSDSSVMVR 139
Db      1389 SAGQDVTGPTGTAFTNAAVLGAQGDIALTGSQANSVQATGCGDGAALNNVTSSTTLTLTA 1448

QY      139 QVGFQGNATA 148
Db      1449 NGNAGNGDAA 1458

RESULT 8
US-10-238-075-1549
; Sequence 1549, Application US/10238075
; Publication No. US20030148324A1
; GENERAL INFORMATION:
; APPLICANT: I.N.S.E.R.M.
; TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isolated
; FILE REFERENCE: BLANDINE
; CURRENT APPLICATION NUMBER: US/10/238,075
; PRIOR FILING DATE: 2002-09-10
; PRIOR APPLICATION NUMBER: 0003145
; NUMBER OF SEQ ID NOS: 1576
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1549
; LENGTH: 182
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-238-075-1549

Query Match      11.8%; Score 89; DB 14; Length 182;
Best Local Similarity 24.5%; Pred. No. 0.17;
Matches 39; Conservative 29; Mismatches 71; Indels 20; Gaps 7;

```

```

QY      1 MLLKVAFAAIVVSGALAGYDQVLT---RVVTHE-MAHASGPDSTLSIYQYGSANA 56
Db      3 IKVIALATFSAVAGSAMA--YDGTITFTGKVVAQCTVNTSDKDLAVL-----PTVAT 56

QY      57 LALQSDARKSETT---ITQSGYNGADVGGQA-----DNSTIELTQNGFRNNATIDOWN 108
Db      57 SSLKNAATSGLTPTFAIRLTGTCATGMSAQNVKAYFPPSSNIDLATHNLKNTAT--PTKA 114

QY      109 KNSDITVQYGNNAALVNQATSDSSVVVRQVGFQGNAT 147
Db      115 DNVQIQLNNGSTILLGEADNGQDVQSETIGSDGSAT 153

RESULT 9
US-09-793-306-146
; Sequence 146, Application US/09793306
; Patent No. US20020098200A1
; GENERAL INFORMATION:
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Skeiky, Yasir
; APPLICANT: Owendale, Pamela
; APPLICANT: Jen, Shyian
; APPLICANT: Lodes, Michael
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compounds and Methods for Diagnosis and Immunotherapy
; FILE REFERENCE: 014058-008740US
; CURRENT APPLICATION NUMBER: US/09/793,306
; CURRENT FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: US 60/185,037
; PRIOR FILING DATE: 2000-02-25
; PRIOR APPLICATION NUMBER: US 60/223,828
; PRIOR FILING DATE: 2000-08-08
; NUMBER OF SEQ ID NOS: 164
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 146
; LENGTH: 597
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: mTCH#3-His
US-09-793-306-146

Query Match      11.7%; Score 88.5; DB 9; Length 597;
Best Local Similarity 35.9%; Pred. No. 1;
Matches 28; Conservative 7; Mismatches 32; Indels 11; Gaps 4;

QY      74 GYGNADVGGADNSSTIELTQNGFRNNATIDQ--MNAKNSDITVQYGNNAALVNQATAS 131
Db      387 GFGNAGDINTGFCNAGD--TNTGFGNAGFFNMIGNAGNEDMGVNGGSGFNVGVGN--AG 442

QY      132 DSSVMVRQVGFQGNATAN 149
Db      443 NQS-----VGFQGNAGTLN 455

RESULT 10
US-10-437-963-162284
; Sequence 162284, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B

```

```
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 162284
; LENGTH: 154
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_61390C.1.pep
US-10-437-963-162284

Query Match      11.4%; Score 86.5; DB 16; Length 154;
Best Local Similarity 23.5%; Pred. No. 0.26;
Matches 40; Conservative 21; Mismatches 56; Indels 53; Gaps 7;

QY 5 KVAFAAIVVSGSALAGYDQVLTVRVTHMAHAGSPDSTLSIYQYGSANAAALQSDAR 64
   ||| |||
   ||| |||
Db 5 KLAALGFVLISGLAS-----AVRVERYSNAQSGTSGEGTGSVNGAGAGLG----- 53
   ||| |||
   ||| |||
QY 65 KSETTITQSGY-----GNGADYVGGQADNSTIELTQNGFRNNA-TI 103
   ||| |||
   ||| |||
Db 54 -SGTSGRSGNGAHASGGGGGGGYQYGGSGSGSGSSQTSQNGYYGSSEAT 112
   ||| |||
   ||| |||
QY 104 DQNAKNSDITV-----GQYGGNNAALVNQTASDSSVMVRQVGFNG 144
   ||| |||
   ||| |||
Db 113 NTWYQYANANAGNGGGGQGYGGGAGV-----GSGS-----GFGD 151
   ||| |||
   ||| |||

RESULT 11
US-09-996-634-131
; Sequence 131, Application US/09996634
; Patent No. US20020172684A1
; GENERAL INFORMATION:
; APPLICANT: Nano, Francis
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
; TITLE OF INVENTION: Immunostimulatory Peptides
; FILE REFERENCE: 61260
; CURRENT APPLICATION NUMBER: US/09/996,634
; CURRENT FILING DATE: 2001-11-28
; PRIOR APPLICATION NUMBER: 09/447,135
; PRIOR FILING DATE: 2000-01-03
; PRIOR APPLICATION NUMBER: 08/990,823
; PRIOR FILING DATE: 1997-12-15
; PRIOR APPLICATION NUMBER: US 96/10375
; PRIOR FILING DATE: 1996-06-14
; PRIOR APPLICATION NUMBER: 60/000,254
; PRIOR FILING DATE: 1995-06-15
; NUMBER OF SEQ ID NOS: 169
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 131
; LENGTH: 943
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-996-634-131

Query Match      11.4%; Score 86.5; DB 9; Length 943;
Best Local Similarity 24.8%; Pred. No. 3.1;
Matches 39; Conservative 22; Mismatches 73; Indels 23; Gaps 6;

QY 7 AAFRAAIVVSG-----SALAGYDQVLTVRVTHMAHAGSPDSTLSIYQYGSANAA 56
   ||| |||
   ||| |||
Db 96 ALFVSLVSNLLGNQAPAIATEAAEQWMAQDVAAFMFGYHAGASAAVSALTFFGQALPT 155
   ||| |||
   ||| |||
QY 57 L-----ALQSDARKSETTITQSGYNGADYVGGQADNSTIELTQNGFRNNATIDQWNAKNSD 112
   ||| |||
   ||| |||
Db 156 VAGGALVSAATAAQVTVFRNLGL-ANVGEQN-----VGNNGVGNFNLGSANIGNN 207
   ||| |||
   ||| |||
QY 113 ITVGQYGGNNAALVNQTASDSSVMVRQVGFNGNATAN 149
   ||| |||
   ||| |||
Db 208 IGSNGIGSSNIGFGN-VGPGGLTAALNNIGFGNTGSNN 243
   ||| |||
   ||| |||

US-09-997-182-131
; Sequence 131, Application US/09997182
; Publication No. US20030049263A1
; GENERAL INFORMATION:
; APPLICANT: Nano, Francis
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
; TITLE OF INVENTION: Immunostimulatory Peptides
; FILE REFERENCE: 61258
; CURRENT APPLICATION NUMBER: US/09/997,182
; CURRENT FILING DATE: 2001-11-28
; PRIOR APPLICATION NUMBER: 09/447,135
; PRIOR FILING DATE: 2000-01-03
; PRIOR APPLICATION NUMBER: 08/990,823
; PRIOR FILING DATE: 1997-12-15
; PRIOR APPLICATION NUMBER: US 96/10375
; PRIOR FILING DATE: 1996-06-14
; PRIOR APPLICATION NUMBER: 60/000,254
; PRIOR FILING DATE: 1995-06-15
; NUMBER OF SEQ ID NOS: 169
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 131
; LENGTH: 943
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-997-182-131

Query Match      11.4%; Score 86.5; DB 10; Length 943;
Best Local Similarity 24.8%; Pred. No. 3.1;
Matches 39; Conservative 22; Mismatches 73; Indels 23; Gaps 6;

QY 7 AAFRAAIVVSG-----SALAGYDQVLTVRVTHMAHAGSPDSTLSIYQYGSANAA 56
   ||| |||
   ||| |||
Db 96 ALFVSLVSNLLGNQAPAIATEAAEQWMAQDVAAFMFGYHAGASAAVSALTFFGQALPT 155
   ||| |||
   ||| |||
QY 57 L-----ALQSDARKSETTITQSGYNGADYVGGQADNSTIELTQNGFRNNATIDQWNAKNSD 112
   ||| |||
   ||| |||
Db 156 VAGGALVSAATAAQVTVFRNLGL-ANVGEQN-----VGNNGVGNFNLGSANIGNN 207
   ||| |||
   ||| |||
QY 113 ITVGQYGGNNAALVNQTASDSSVMVRQVGFNGNATAN 149
   ||| |||
   ||| |||
Db 208 IGSNGIGSSNIGFGN-VGPGGLTAALNNIGFGNTGSNN 243
   ||| |||
   ||| |||

US-09-997-181-131
; Sequence 131, Application US/09997181
; Publication No. US20030049269A1
; GENERAL INFORMATION:
; APPLICANT: Nano, Francis
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
; TITLE OF INVENTION: Immunostimulatory Peptides
; FILE REFERENCE: 61257
; CURRENT APPLICATION NUMBER: US/09/997,181
; CURRENT FILING DATE: 2001-11-28
; PRIOR APPLICATION NUMBER: 09/447,135
; PRIOR FILING DATE: 2000-01-03
; PRIOR APPLICATION NUMBER: 08/990,823
; PRIOR FILING DATE: 1997-12-15
; PRIOR APPLICATION NUMBER: US 96/10375
; PRIOR FILING DATE: 1996-06-14
; PRIOR APPLICATION NUMBER: 60/000,254
; PRIOR FILING DATE: 1995-06-15
; NUMBER OF SEQ ID NOS: 169
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 131
; LENGTH: 943
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-997-181-131

Query Match      11.4%; Score 86.5; DB 10; Length 943;
Best Local Similarity 24.8%; Pred. No. 3.1;
Matches 39; Conservative 22; Mismatches 73; Indels 23; Gaps 6;

QY 7 AAFRAAIVVSG-----SALAGYDQVLTVRVTHMAHAGSPDSTLSIYQYGSANAA 56
   ||| |||
   ||| |||
Db 96 ALFVSLVSNLLGNQAPAIATEAAEQWMAQDVAAFMFGYHAGASAAVSALTFFGQALPT 155
   ||| |||
   ||| |||
QY 57 L-----ALQSDARKSETTITQSGYNGADYVGGQADNSTIELTQNGFRNNATIDQWNAKNSD 112
   ||| |||
   ||| |||
Db 156 VAGGALVSAATAAQVTVFRNLGL-ANVGEQN-----VGNNGVGNFNLGSANIGNN 207
   ||| |||
   ||| |||
QY 113 ITVGQYGGNNAALVNQTASDSSVMVRQVGFNGNATAN 149
   ||| |||
   ||| |||
Db 208 IGSNGIGSSNIGFGN-VGPGGLTAALNNIGFGNTGSNN 243
   ||| |||
   ||| |||

US-09-996-634-131
; Sequence 131, Application US/09996634
; Patent No. US20020172684A1
; GENERAL INFORMATION:
; APPLICANT: Nano, Francis
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
; TITLE OF INVENTION: Immunostimulatory Peptides
; FILE REFERENCE: 61260
; CURRENT APPLICATION NUMBER: US/09/996,634
; CURRENT FILING DATE: 2001-11-28
; PRIOR APPLICATION NUMBER: 09/447,135
; PRIOR FILING DATE: 2000-01-03
; PRIOR APPLICATION NUMBER: 08/990,823
; PRIOR FILING DATE: 1997-12-15
; PRIOR APPLICATION NUMBER: US 96/10375
; PRIOR FILING DATE: 1996-06-14
; PRIOR APPLICATION NUMBER: 60/000,254
; PRIOR FILING DATE: 1995-06-15
; NUMBER OF SEQ ID NOS: 169
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 131
; LENGTH: 943
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-996-634-131

Query Match      11.4%; Score 86.5; DB 9; Length 943;
Best Local Similarity 24.8%; Pred. No. 3.1;
Matches 39; Conservative 22; Mismatches 73; Indels 23; Gaps 6;

QY 7 AAFRAAIVVSG-----SALAGYDQVLTVRVTHMAHAGSPDSTLSIYQYGSANAA 56
   ||| |||
   ||| |||
Db 96 ALFVSLVSNLLGNQAPAIATEAAEQWMAQDVAAFMFGYHAGASAAVSALTFFGQALPT 155
   ||| |||
   ||| |||
QY 57 L-----ALQSDARKSETTITQSGYNGADYVGGQADNSTIELTQNGFRNNATIDQWNAKNSD 112
   ||| |||
   ||| |||
Db 156 VAGGALVSAATAAQVTVFRNLGL-ANVGEQN-----VGNNGVGNFNLGSANIGNN 207
   ||| |||
   ||| |||
QY 113 ITVGQYGGNNAALVNQTASDSSVMVRQVGFNGNATAN 149
   ||| |||
   ||| |||
Db 208 IGSNGIGSSNIGFGN-VGPGGLTAALNNIGFGNTGSNN 243
   ||| |||
   ||| |||

US-09-997-181-131
; Sequence 131, Application US/09997181
; Publication No. US20030049269A1
; GENERAL INFORMATION:
; APPLICANT: Nano, Francis
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
; TITLE OF INVENTION: Immunostimulatory Peptides
; FILE REFERENCE: 61257
; CURRENT APPLICATION NUMBER: US/09/997,181
; CURRENT FILING DATE: 2001-11-28
; PRIOR APPLICATION NUMBER: 09/447,135
; PRIOR FILING DATE: 2000-01-03
; PRIOR APPLICATION NUMBER: 08/990,823
; PRIOR FILING DATE: 1997-12-15
; PRIOR APPLICATION NUMBER: US 96/10375
; PRIOR FILING DATE: 1996-06-14
; PRIOR APPLICATION NUMBER: 60/000,254
; PRIOR FILING DATE: 1995-06-15
; NUMBER OF SEQ ID NOS: 169
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 131
; LENGTH: 943
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-997-181-131

Query Match      11.4%; Score 86.5; DB 10; Length 943;
Best Local Similarity 24.8%; Pred. No. 3.1;
Matches 39; Conservative 22; Mismatches 73; Indels 23; Gaps 6;
```

Qy	7	A	A	F	A	A	I	V	T	S	G	-----	S	A	L	A	G	V	Y	D	O	L	V	T	R	V	T	H	E	M	A	H	A	S	G	P	D	S	T	L	S	-	I	Y	O	G	S	N	A	A	56	
Db	96	A	L	F	V	S	V	S	N	I	L	Q	W	A	P	A	I	E	A	E	Y	E	Q	W	A	D	V	A	M	F	G	H	A	G	A	S	A	V	S	A	L	T	F	F	G	A	L	P	T	155		
Qy	57	L	-----	A	L	O	S	D	A	R	K	S	T	T	T	I	Q	S	Y	G	N	A	D	V	G	G	A	D	S	T	I	E	L	T	Q	N	G	F	R	N	N	A	T	I	D	O	N	N	A	S	D	112
Db	156	V	A	G	G	A	L	V	S	A	A	A	A	A	Q	V	T	T	R	F	R	N	L	G	L	-	A	N	V	G	E	N	-----	V	G	N	G	V	N	F	N	L	G	S	A	N	T	G	N	207		
Qy	113	I	T	Y	Q	V	Q	G	G	N	A	A	L	V	N	Q	T	A	S	D	S	S	V	M	V	R	Q	V	G	F	N	N	A	T	A	N	149															
Db	208	I	G	S	G	N	I	G	S	N	I	G	F	G	-	V	G	P	L	T	A	A	L	N	I	G	F	N	G	T	S	N	243																			

```

RESULT 14
US-10-282-122A-64369
; Sequence 64369, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsui
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 64369
; LENGTH: 3300
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-10-282-122A-64369

```

Db	175	VAGGALYSAAAQVTFVFNGL-ANVGE	-----VNGNGVGNFLGSANTGN	226
Qy	113	ITVQYGGNNALVNQTSDDSVVRQVGF	NNATAN	149
		:	:	
Db	227	IGSGNIGSNTGFGN-VGPGHTAALNNIG	FNTGTSNN	262

RESULT 15
 US-10-383-930-38
 ; Sequence 38, Application US/10383930
 ; Publication No. US20040127400A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Smith, Daniel J
 ; APPLICANT: Taubman, Martin A
 ; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
 ; FILE REFERENCE: 25669-018
 ; CURRENT APPLICATION NUMBER: US/10/383,930
 ; CURRENT FILING DATE: 2003-03-07
 ; PRIOR APPLICATION NUMBER: 60/402,483
 ; PRIOR FILING DATE: 2002-08-08
 ; PRIOR APPLICATION NUMBER: 60/363,209
 ; PRIOR FILING DATE: 2002-03-07
 ; NUMBER OF SEQ ID NOS: 41
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 38
 ; LENGTH: 1554
 ; TYPE: PRT
 ; ORGANISM: Streptococcus sobrinus
 US-10-383-930-38

Search completed: August 2, 2004, 15:36:10
Job time : 37.8 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:48:33 ; Search time 167.9 Seconds
(without alignments)
877.809 Million cell updates/sec

Title: US-09-543-407-16

Perfect score: 757

Sequence: 1 MKLLKYAFRAIVVSGSALA.....DSSVMVRQVFGGNATANQV 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 6019581 seqs, 976053577 residues

Total number of hits satisfying chosen parameters: 6019581

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending_Parents_AA_Main:*

- 1: /cgn2_6/ptodata/2/paa/PCTUS_COMB.pcp.*
- 2: /cgn2_6/ptodata/2/paa/US06_COMB.pcp.*
- 3: /cgn2_6/ptodata/2/paa/US07_COMB.pcp.*
- 4: /cgn2_6/ptodata/2/paa/US08_COMB.pcp.*
- 5: /cgn2_6/ptodata/2/paa/US081_COMB.pcp.*
- 6: /cgn2_6/ptodata/2/paa/US082_COMB.pcp.*
- 7: /cgn2_6/ptodata/2/paa/US083_COMB.pcp.*
- 8: /cgn2_6/ptodata/2/paa/US084_COMB.pcp.*
- 9: /cgn2_6/ptodata/2/paa/US085_COMB.pcp.*
- 10: /cgn2_6/ptodata/2/paa/US086_COMB.pcp.*
- 11: /cgn2_6/ptodata/2/paa/US087_COMB.pcp.*
- 12: /cgn2_6/ptodata/2/paa/US088_COMB.pcp.*
- 13: /cgn2_6/ptodata/2/paa/US089_COMB.pcp.*
- 14: /cgn2_6/ptodata/2/paa/US090_COMB.pcp.*
- 15: /cgn2_6/ptodata/2/paa/US091_COMB.pcp.*
- 16: /cgn2_6/ptodata/2/paa/US092_COMB.pcp.*
- 17: /cgn2_6/ptodata/2/paa/US093_COMB.pcp.*
- 18: /cgn2_6/ptodata/2/paa/US094_COMB.pcp.*
- 19: /cgn2_6/ptodata/2/paa/US095_COMB.pcp.*
- 20: /cgn2_6/ptodata/2/paa/US096_COMB.pcp.*
- 21: /cgn2_6/ptodata/2/paa/US097A_COMB.pcp.*
- 22: /cgn2_6/ptodata/2/paa/US097B_COMB.pcp.*
- 23: /cgn2_6/ptodata/2/paa/US098_COMB.pcp.*
- 24: /cgn2_6/ptodata/2/paa/US099A_COMB.pcp.*
- 25: /cgn2_6/ptodata/2/paa/US099B_COMB.pcp.*
- 26: /cgn2_6/ptodata/2/paa/US100_COMB.pcp.*
- 27: /cgn2_6/ptodata/2/paa/US101_COMB.pcp.*
- 28: /cgn2_6/ptodata/2/paa/US102_COMB.pcp.*
- 29: /cgn2_6/ptodata/2/paa/US103_COMB.pcp.*
- 30: /cgn2_6/ptodata/2/paa/US104_COMB.pcp.*
- 31: /cgn2_6/ptodata/2/paa/US106_COMB.pcp.*
- 32: /cgn2_6/ptodata/2/paa/US107_COMB.pcp.*
- 33: /cgn2_6/ptodata/2/paa/US108_COMB.pcp.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
------------	-------	-------------	--------	-------	-------------

1	757	100.0	151	19	US-09-543-407-16	Sequence 16, Appl
2	658	86.9	151	19	US-09-543-407-5	Sequence 5, Appl
3	653	86.3	151	6	US-08-233-642A-57	Sequence 57, Appl
4	618	81.6	151	19	US-09-543-407-18	Sequence 18, Appl
5	580	76.6	151	19	US-09-543-407-12	Sequence 12, Appl
6	579	76.5	131	19	US-09-543-407-31	Sequence 31, Appl
7	578	76.4	151	19	US-09-543-407-14	Sequence 14, Appl
8	577	76.2	151	19	US-09-543-407-26	Sequence 26, Appl
9	575	76.0	151	19	US-09-543-407-24	Sequence 24, Appl
10	574	75.8	151	19	US-09-543-407-20	Sequence 20, Appl
11	568	75.0	151	19	US-09-543-407-28	Sequence 28, Appl
12	567	74.9	151	19	US-09-543-407-22	Sequence 22, Appl
13	566	74.8	151	19	US-09-543-407-30	Sequence 30, Appl
14	554	73.2	120	6	US-08-233-642A-55	Sequence 34, Appl
15	487	64.3	120	6	US-08-233-642A-55	Sequence 55, Appl
16	484	63.9	151	19	US-09-543-407-7	Sequence 7, Appl
17	481	63.5	151	13	US-08-978-878-4	Sequence 4, Appl
18	479	63.3	151	21	US-08-741-873B-4	Sequence 2, Appl
19	479	63.3	151	33	US-60-352-946-2	Sequence 2, Appl
20	479	63.3	151	33	US-60-444-371-2	Sequence 2, Appl
21	468	61.8	158	16	US-09-252-691-5834	Sequence 5834, Ap
22	468	61.8	158	16	US-09-252-691C-5834	Sequence 5834, Ap
23	468	61.8	158	30	US-10-417-886-5834	Sequence 5834, Ap
24	414	54.7	131	13	US-08-978-878-2	Sequence 2, Appl
25	414	54.7	131	21	US-09-741-873B-2	Sequence 2, Appl
26	401	53.0	109	19	US-09-543-407-35	Sequence 35, Appl
27	347	45.8	68	19	US-09-543-407-37	Sequence 37, Appl
28	250.5	33.1	70	19	US-09-543-407-32	Sequence 32, Appl
29	237	31.3	48	19	US-09-543-407-39	Sequence 39, Appl
30	122.5	16.2	186	16	US-09-252-691-5833	Sequence 5833, Ap
31	122.5	16.2	186	16	US-09-252-691C-5833	Sequence 5833, Ap
32	122.5	16.2	186	30	US-10-417-886-5833	Sequence 5833, Ap
33	118.5	15.7	151	19	US-09-543-407-8	Sequence 8, Appl
34	113.5	15.0	445	29	US-10-369-493-20638	Sequence 20638, A
35	113.5	15.0	445	33	US-60-360-039-20638	Sequence 20638, A
36	111.5	14.7	151	19	US-09-543-407-6	Sequence 6, Appl
37	101.5	13.4	590	3	US-07-965-673A-10	Sequence 10, Appl
38	101.5	13.4	590	22	US-09-791-537-29229	Sequence 29229, A
39	101.5	13.4	590	22	US-09-791-537-111372	Sequence 111372, A
40	101	13.3	599	22	US-09-791-537-29228	Sequence 29228, A
41	101	13.3	599	22	US-09-791-537-111371	Sequence 111371, A
42	99.5	13.1	582	22	US-09-791-537-94843	Sequence 94843, A
43	99.5	13.1	646	22	US-09-791-537-30765	Sequence 30765, A
44	99.5	13.1	646	22	US-09-791-537-32862	Sequence 32862, A
45	98	12.9	145	21	US-09-739-449-8854	Sequence 8854, Ap

ALIGNMENTS

RESULT 1
US-09-543-407-16
; Sequence 16, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.

US-09-543-407-16

Query Match 100.0%; Score 757; DB 19; Length 151;
Best Local Similarity 100.0%; Pred. No. 3.4e-76;
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIVVSGSALAGVVDLVTRVVTHEMAHASGPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAAPAAIVVSGSALAGVVDLVTRVVTHEMAHASGPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
QY 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151
DB 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151

RESULT 2

US-09-543-407-5

; Sequence 5, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Salmonella enteritidis
US-09-543-407-5

Query Match 86.9%; Score 658; DB 19; Length 151;
Best Local Similarity 90.1%; Pred. No. 4.5e-65;
Matches 136; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIVVSGSALAGVVDLVTRVVTHEMAHASGPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAAPAAIVVSGSALAGVVDLVTRVVTHEMAHASGPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
QY 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151
DB 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151

RESULT 3

US-08-233-642A-57

; Sequence 57, Application US/08233642A
; GENERAL INFORMATION:
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Clouthier, Sharon C.
; APPLICANT: Doran, James L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
; TITLE OF INVENTION: BASED VACCINES
; NUMBER OF SEQUENCES: 58
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle

; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,642A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 151 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-233-642A-57

Query Match 86.3%; Score 653; DB 6; Length 151;
Best Local Similarity 89.4%; Pred. No. 1.6e-64;
Matches 135; Conservative 3; Mismatches 13; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIVVSGSALAGVVDLVTRVVTHEMAHASGPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAAPAAIVVSGSALAGVVDLVTRVVTHEMAHASGPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
QY 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151
DB 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151

RESULT 4

US-09-543-407-18

; Sequence 18, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-18

Query Match 81.6%; Score 618; DB 19; Length 151;
Best Local Similarity 76.6%; Pred. No. 1.4e-60;
Matches 131; Conservative 0; Mismatches 0; Indels 40; Gaps 2;

```
QY 1 MKLLKVAFAAIVVSGSALAGV-----YDQLVTRVVTHEMAHASG 40
Db 1 MKLLKVAFAAIVVSGSALAGVPCMGGGGHNHGGSSGPDYDQLVTRVVTHEMAH--- 57
QY 41 PDSTLSIYQGSANAALALQSDARKSETTITQSGYGNAGADVQGGADNSTIETQNGFRNN 100
Db 58 -----ALQSDARKSETTITQSGYGNAGADVQGGADNSTIETQNGFRNN 100
QY 101 ATIDOWNAKNSDITVQYGGNNAALVNCATSDSSVMVRQVGFGNATANQY 151
Db 101 ATIDOWNAKNSDITVQYGGNNAALVNCATSDSSVMVRQVGFGNATANQY 151

RESULT 5
US-09-543-407-12
; Sequence 12, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 151
; TYPE: PR1
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-12

Query Match 76.6%; Score 580; DB 19; Length 151;
Best Local Similarity 80.1%; Pred. No. 2.6e-56;
Matches 121; Conservative 7; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVYDQLVTRVVTHEMAHASGPDSTLSIYQGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVPCMGGGGHNHGGSSGPDSTLSIYQGSANAALALQ 60
QY 61 SDARKSETTITQSGYGNAGADVQGGADNSTIETQNGFRNNATIDOWNAKNSDITVQYGG 120
Db 61 SDARKSETTITQSGYGNAGADVQGGADNSTIETQNGFRNNATIDOWNAKNSDITVQYGG 120
QY 121 NNAALVNCATSDSSVMVRQVGFGNATANQY 151
Db 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151

RESULT 6
US-09-543-407-31
; Sequence 31, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 131

QY 1 MKLLKVAFAAIVVSGSALAGVYDQLVTRVVTHEMAHASGPDSTLSIYQGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVPCMGGGGHNHGGSSGPDSTLSIYQGSANAALALQ 60
QY 61 SDARKSETTITQSGYGNAGADVQGGADNSTIETQNGFRNNATIDOWNAKNSDITVQYGG 120
Db 61 SDARKSETTITQSGYGNAGADVQGGADNSTIETQNGFRNNATIDOWNAKNSDITVQYGG 120
QY 121 NNAALVNCATSDSSVMVRQVGFGNATANQY 151
Db 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151

RESULT 7
US-09-543-407-14
; Sequence 14, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 151
; TYPE: PR1
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-14

Query Match 76.4%; Score 578; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 4.4e-56;
Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVYDQLVTRVVTHEMAHASGPDSTLSIYQGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVPCMGGGGHNHGGSSGPDSTLSIYQGSANAALALQ 60
QY 61 SDARKSETTITQSGYGNAGADVQGGADNSTIETQNGFRNNATIDOWNAKNSDITVQYGG 120
Db 61 SDARKSETTITQSGYGNAGADVQGGADNSTIETQNGFRNNATIDOWNAKNSDITVQYGG 120
QY 121 NNAALVNCATSDSSVMVRQVGFGNATANQY 151
Db 121 LVTRVVTHEMAHASVMVRQVGFGNATANQY 151

RESULT 8
US-09-543-407-26
; Sequence 26, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
```

```
; TYPE: PR1
; ORGANISM: Salmonella enteritidis
US-09-543-407-31

Query Match 76.5%; Score 579; DB 19; Length 131;
Best Local Similarity 98.3%; Pred. No. 2.8e-56;
Matches 113; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 37 HASGPDSTLSIYQGSANAALALQSDARKSETTITQSGYGNAGADVQGGADNSTIETQNG 96
Db 17 NSGSPDSTLSIYQGSANAALALQSDARKSETTITQSGYGNAGADVQGGADNSTIETQNG 76
QY 97 FRNNATIDOWNAKNSDITVQYGGNNAALVNCATSDSSVMVRQVGFGNATANQY 151
Db 77 FRNNATIDOWNAKNSDITVQYGGNNAALVNCATSDSSVMVRQVGFGNATANQY 131

RESULT 7
US-09-543-407-14
; Sequence 14, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 151
; TYPE: PR1
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-14

Query Match 76.4%; Score 578; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 4.4e-56;
Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVYDQLVTRVVTHEMAHASGPDSTLSIYQGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVPCMGGGGHNHGGSSGPDSTLSIYQGSANAALALQ 60
QY 61 SDARKSETTITQSGYGNAGADVQGGADNSTIETQNGFRNNATIDOWNAKNSDITVQYGG 120
Db 61 SDARKSETTITQSGYGNAGADVQGGADNSTIETQNGFRNNATIDOWNAKNSDITVQYGG 120
QY 121 NNAALVNCATSDSSVMVRQVGFGNATANQY 151
Db 121 LVTRVVTHEMAHASVMVRQVGFGNATANQY 151

RESULT 8
US-09-543-407-26
; Sequence 26, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
```

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-26

Query Match          76.2%; Score 577; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 5.7e-56;
Matches 122; Conservative 5; Mismatches 24; Indels 0; Gaps 0;

Qy 1 MKLLKVAFAAIVVSGSALAGVVDQLVTRVVTHEMAHASGPDSTLSIYQYGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNHNGSGSPDSTLSIYQYGSANAALYDQ 60
Qy 61 SDARKSETTITQSGYNGADVCGADNSTIELTQNGFRNNATIDOWNAKNSDITVGYGG 120
Db 61 LVTRVVTHEMAHAGYNGADVCGADNSTIELTQNGFRNNATIDOWNAKNSDITVGYGG 120
Qy 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 9
US-09-543-407-24
; Sequence 24, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-24

Query Match          76.0%; Score 575; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 9.5e-56;
Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;

Qy 1 MKLLKVAFAAIVVSGSALAGVVDQLVTRVVTHEMAHASGPDSTLSIYQYGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNHNGSGSPDSTLSIYQYGSANAALALQ 60
Qy 61 SDARKSETTITQSGYNGADVCGADNSTIELTQNGFRNNATIDOWNAKNSDITVGYGG 120
Db 61 SDARKSETTITQSGYNGADVCGADNSTIELTQNGFRNNATIDOWNAKNSDITVGYGG 120
Qy 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
Db 121 HEWAHANQTASDSSVMVRQVGFNNATANQY 151

RESULT 10
US-09-543-407-20
; Sequence 20, Application US/09543407
```

```
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-20

Query Match          75.8%; Score 574; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 1.2e-55;
Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;

Qy 1 MKLLKVAFAAIVVSGSALAGVVDQLVTRVVTHEMAHASGPDSTLSIYQYGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNHNGSGSPDSTLSIYQYGSANAALALQ 60
Qy 61 SDARKSETTITQSGYNGADVCGADNSTIELTQNGFRNNATIDOWNAKNSDITVGYGG 120
Db 61 SDARKYDQLVTRVVTHEMAHAGQADNSTIELTQNGFRNNATIDOWNAKNSDITVGYGG 120
Qy 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 11
US-09-543-407-28
; Sequence 28, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-28

Query Match          75.0%; Score 568; DB 19; Length 151;
Best Local Similarity 80.1%; Pred. No. 5.8e-55;
Matches 121; Conservative 4; Mismatches 26; Indels 0; Gaps 0;

Qy 1 MKLLKVAFAAIVVSGSALAGVVDQLVTRVVTHEMAHASGPDSTLSIYQYGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNHNGSGSPDSTLSIYQYGSANAALALQ 60
Qy 61 SDARKSETTITQSGYNGADVCGADNSTIELTQNGFRNNATIDOWNAKNSDITVGYGG 120
```


Db 61 SDARKSETTITQSGYGNADYDQLVTRVVTTHMAHAFNNATIDQWNAKNSDIIVGOYGG 120
QY 121 NNAALVNQTASDSSVMVRQVFGNNATANQY 151
Db 121 NNAALVNQTASDSSVMVRQVFGNNATANQY 151

RESULT 12
US-09-543-407-22
; Sequence 22, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 22
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-22

Query Match 74.9%; Score 567; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 7.5e-55;
Matches 122; Conservative 5; Mismatches 24; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVVSGSALAGYDQLVTRVVTTHMAHAGSPDSTLSIYQGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGYDQLVTRVVTTHMAHAGSPDSTLSIYQGSANAALALQ 60
QY 61 SDARKSETTITQSGYGNADYDQGGADNSTIELTQNGFRNNATIDQWNAKNSDIIVGOYGG 120
Db 61 SDARKSETTITQSGYGNADYDQGGADNDYDQLVTRVVTTHMAHADQWNAKNSDIIVGOYGG 120
QY 121 NNAALVNQTASDSSVMVRQVFGNNATANQY 151
Db 121 NNAALVNQTASDSSVMVRQVFGNNATANQY 151

RESULT 13
US-09-543-407-30
; Sequence 30, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.

US-09-543-407-30

Query Match 74.8%; Score 566; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 9.7e-55;
Matches 122; Conservative 5; Mismatches 24; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVVSGSALAGYDQLVTRVVTTHMAHAGSPDSTLSIYQGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGYDQGGGNGHNGSGSPDSTLSIYQGSANAALALQ 60
QY 61 SDARKSETTITQSGYGNADYDQGGADNSTIELTQNGFRNNATIDQWNAKNSDIIVGOYGG 120
Db 61 SDARKSETTITQSGYGNADYDQGGADNSTIELTQNGFRNNATYDQLVTRVVTTHMAHAGG 120
QY 121 NNAALVNQTASDSSVMVRQVFGNNATANQY 151
Db 121 NNAALVNQTASDSSVMVRQVFGNNATANQY 151

RESULT 14
US-09-543-407-34
; Sequence 34, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 109
; TYPE: PRT
; ORGANISM: Salmonella enteritidis
US-09-543-407-34

Query Match 73.2%; Score 554; DB 19; Length 109;
Best Local Similarity 100.0%; Pred. No. 1.4e-53;
Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 43 STLSIYQGSANAALALQSDARKSETTITQSGYGNADYDQGGADNSTIELTQNGFRNNAT 102
Db 1 STLSIYQGSANAALALQSDARKSETTITQSGYGNADYDQGGADNSTIELTQNGFRNNAT 60
QY 103 IDQWNAKNSDIIVGOYGGNNAALVNQTASDSSVMVRQVFGNNATANQY 151
Db 61 IDQWNAKNSDIIVGOYGGNNAALVNQTASDSSVMVRQVFGNNATANQY 109

RESULT 15
US-08-233-642A-55
; Sequence 55, Application US/08233642A
; GENERAL INFORMATION:
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Clouthier, Sharon C.
; APPLICANT: Doran, James L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
; BASED VACCINES
; NUMBER OF SEQUENCES: 58
; NUMBER OF SEQUENCES: -
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:

```
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,642A
; FILING DATE: 28-APR-1994
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 55:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 120 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-233-642A-55

Query Match          64.3%; Score 487; DB 6; Length 120;
Best Local Similarity 97.9%; Pred. No. 5.2e-46;
Matches 95; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 37 HASGPDSTLSIYQGSANAALALQSDARKSETTITQSGYNGADVQGGADNSTIELTQNG 96
   :|||||
Db 16 NSSGPDSTLSIYQGSANAALALQSDARKSETTITQSGYNGADVQGGADNSTIELTQNG 75
   :|||||

QY 97 FRNNATIDOWNAKNSDITVGYGGNNAALVNQTSDS 133
   :|||||
Db 76 FRNNATIDOWNAKNSDITVGYGGNNAALVNQTSDS 112
   :|||||

Search completed: August 2, 2004, 15:26:43
Job time : 168.9 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:49:38 ; Search time 17.8 Seconds
(without alignments)
888.146 Million cell updates/sec

Title: US-09-543-407-16
Perfect score: 757
Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVWVQVGFNNATANQY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 601315 seqs, 104695340 residues

Total number of hits satisfying chosen parameters: 601315

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents AA New:*

1: /cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pep:*
2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep:*
3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep:*
4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep:*
5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:*
6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:*
7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	481	63.5	151	US-09-741-873C-4	Sequence 4, Appl
2	414	54.7	131	US-09-741-873C-2	Sequence 2, Appl
3	93	12.3	348	US-10-004-115B-34	Sequence 34, Appl
4	84	11.1	385	US-60-556-841-1453	Sequence 1453, Ap
5	82.5	10.9	597	US-10-784-592-35	Sequence 35, Appl
6	82	10.8	1236	US-10-873-528-109	Sequence 109, App
7	81.5	10.8	1158	US-10-170-205E-32312	Sequence 32312, A
8	81.5	10.8	1158	US-10-501-035-366	Sequence 366, App
9	80.5	10.6	1095	US-09-806-709-69	Sequence 69, Appl
10	80.5	10.6	1536	US-09-806-709-67	Sequence 67, Appl
11	80	10.6	485	US-10-170-205E-11455	Sequence 11455, A
12	80	10.6	503	US-10-170-205E-11130	Sequence 11130, A
13	80	10.6	503	US-10-170-205E-11456	Sequence 11456, A
14	80	10.6	546	US-10-425-115-362263	Sequence 362263, A
15	80	10.6	829	US-60-579-902-7335	Sequence 7335, Ap
16	79	10.4	4326	PCT-US04-12717-18	Sequence 18, Appl
17	79	10.4	4326	US-10-831-070-18	Sequence 18, Appl
18	78.5	10.4	358	US-10-767-701-40444	Sequence 40444, A
19	77	10.2	841	US-60-565-632-7906	Sequence 7906, Ap
20	77	10.2	841	US-60-579-062-7906	Sequence 7906, Ap
21	77	10.2	10203	US-10-724-972A-4098	Sequence 4098, Ap
22	76.5	10.1	388	US-10-805-394-5044	Sequence 5044, Ap
23	76.5	10.1	630	US-60-581-351-9806	Sequence 9806, Ap
24	76.5	10.1	631	US-60-581-351-9807	Sequence 9807, Ap
25	76.5	10.1	892	US-09-952-267B-5	Sequence 5, Appl
26	76.5	10.1	892	US-10-872-768-5	Sequence 5, Appl

ALIGNMENTS

RESULT 1

US-09-741-873C-4
; Sequence 4: Application US/09741873C

; GENERAL INFORMATION: Staffan
; APPLICANT: Normark, Staffan
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli

US-09-741-873C-4

Query Match 63.5%; Score 481; DB 5; Length 151;
Best Local Similarity 66.2%; Pred. No. 1.4e-39;
Matches 100; Conservative 19; Mismatches 32; Indels 0; Gaps 0;

QY	1	MLLLKVAFAAIVVSGSALA	YVDQLVTRVTHMAHAGSPDSTLSIYQVGSNALALQ	60
Db	1	MLLLKVAFAAIVVSGSALA	VGVVQVGGNGGNGNNGPNSLNLYQVGGNSALALQ	60
QY	61	SDARKSETTITGSGYNGADVGQCADNSTIELTQNGFRNNATIDOWNAKNSDITVQYQYG	120	
Db	61	TDARNSDLITTHGGGGGADVGQSDSSIDITQRCFGNSATLDQWNGKNSMTVKQFG	120	
QY	121	NNAALVNQASDSSVWVQVGFNNATANQY	151	
Db	121	NGCAAVDQTASNSVNVTVQVGFNNATANQY	151	


```

/ NAME/KEY: MISC FEATURE
/ LOCATION: (139)..(139)
/ OTHER INFORMATION: putative copper binding site
/ FEATURE:
/ NAME/KEY: MISC FEATURE
/ LOCATION: (141)..(141)
/ OTHER INFORMATION: putative copper binding site
/ FEATURE:
/ NAME/KEY: MISC FEATURE
/ LOCATION: (181)..(181)
/ OTHER INFORMATION: putative copper binding site
/ FEATURE:
/ NAME/KEY: MISC FEATURE
/ LOCATION: (183)..(183)
/ OTHER INFORMATION: putative copper binding site
/ FEATURE:
/ NAME/KEY: MISC FEATURE
/ LOCATION: (514)..(514)
/ OTHER INFORMATION: putative copper binding site
/ FEATURE:
/ NAME/KEY: MISC FEATURE
/ LOCATION: (566)..(566)
/ OTHER INFORMATION: putative copper binding site
US-10-784-592-35

```

```

Query Match      10.9%; Score 82.5; DB 6; Length 597;
Best Local Similarity 26.1%; Pred.No. 9;
Matches         46; Conservative 27; Mismatches 50; Indels 53; Gaps 11;

QY       7 AAFAAIIVVSSALAGVYDQLVTRVVTHMAHAGSPDSLTSIIYOGSNAALAQSDAR--  64
          ||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Db      185 ASPEEV---GLGYGAFVVLPKRAV-----HPADRDVTLVHEWPTASTAQTWMANLKAG 236

QY       65 -----KSEITT-----ITSGYGNG-----AD-VGGGADNST-----IELTONGF  97
          :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db     237 NLGFSAKGESAGMGGMQMCGDMNGMVGGAADGTGGGGSASDIAHVLPGLPQLNGF  296

QY       98 RNNATIDQWAKNSDITTVGOYGGGNNAALVNQTTASDS-----VVNVQVFQFN  144
          :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db    297 --SPTANDWAA--LDENAGMY---DAFTVNQNASGTTLLPAKPQOLVRLRVINSGN  345

```

```

RESULT 6
US-10-873-528-109
; Sequence 109, Application US/10873528
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics Limited
; APPLICANT: Gilbert, Christophe FG
; APPLICANT: Hansbro, Philip M
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21129WO
; CURRENT APPLICATION NUMBER: US/10/873,528
; CURRENT FILING DATE: 2004-06-23
; PRIOR APPLICATION NUMBER: US/09/769,787
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: GB 9816337.1
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: US 60/125164
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 388
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 109
; LENGTH: 1236
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-873-528-109

```

```

Query Match      10.8%; Score 82; DB 6; Length 1236;
Best Local Similarity 26.0%; Pred. No. 25;
Matches 39; Conservative 16; Mismatches 61; Indels 34; Gaps 7;

Qy      16  GSAALGYDQIVTRVVTNHAHSGPDSITISYQG-SANAALALQSDARK----- 65
      |||  : : :

```

```

Db      235 GSKLFTVTIVYVNPKTNDLGNISMPRGVSIYNSGTSTQTMLTLGSDLGKPSGVKNYIT 20
Qy      66 -----SETTITO-SGY--GNGADV-----GQGADNS-TIELTQNGFRNNATI 103
Db      295 DKNRGVLISYNTSTMTTQSGYTWGCAQMGNGFFAKGKGYLGTSSWTVPIT--GTDTSFTF 352
Qy      104 DOWNAKNSDITVQYQGGNNAALVNQTSDS 133
Db      353 TTYAARTDRIGINYFGGKGVESSTTSQS 382

RESULT 7
US-10-170-205E-32312
; Sequence 32312, Application US/10170205E
; GENERAL INFORMATION:
; APPLICANT: ADAMS, Mark
; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS ORNS
; TITLE OF INVENTION: CAPTURE AGENTS, AND USES THEREOF
; FILE REFERENCE: CL001381
; CURRENT APPLICATION NUMBER: US/10/170,205E
; CURRENT FILING DATE: 2002-06-13
; NUMBER OF SEQ ID NOS: 40312
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 32312
; LENGTH: 1158
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-170-205E-32312

Query Match 10.8%; Score 81.5; DB 6; Length 1158;
Best Local Similarity 21.1%; Pred. No. 26;
Matches 30; Conservative 28; Mismatches 67; Indels 17; Gaps 3

Qy      5 KVAATAFIIVVGSSALAGYVDLTVRVVTHMAHASGPDSTLSTIYQYGSAUALALQSDAR 64
Db      297 RIAAAMARIQENGKQDVGSY----QLPKGMSSHNGQARTSSSKLMASKSSATVFNPMG 351
Qy      65 KSETTITQSGYGNGAD--VGQGADNS-----TIELTQNGFRNNATIDOWNAKNSD 112
Db      352 ITPSGPKRGYKNSLERNNIQAMNNLLLLHLKSOIIPKPMNGHSHSRGSI FESSTP 411
Qy      113 ITVQYQGGNNAALVNQTSADS 134
Db      412 TTIDVSDNNPSPFTDSDGDS 433

```

```

RESULT 8
US-10-501-035-366
; Sequence 366, Application US/10501035
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: IDENTIFICATION OF POLYNUCLEOTIDES AND POLYPEPTIDE FOR PREDICTING
; TITLE OF INVENTION: ACTIVITY OF COMPOUNDS THAT INTERACT WITH PROTEIN TYROSINE KINASE
; TITLE OF INVENTION: AND/OR PROTEIN TYROSINE KINASE PATHWAYS
; FILE REFERENCE: D0185 PCT
; CURRENT APPLICATION NUMBER: US/10/501.035
; CURRENT FILING DATE: 2004-07-09
; PRIOR APPLICATION NUMBER: US 60/350,061
; PRIOR FILING DATE: 2002-01-18
; NUMBER OF SEQ ID NOS: 795
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 366
; LENGTH: 1158
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-501-035-366

```

```
Query Match      10.8%; Score 81.5; DB 6; Length 1158;
Best Local Similarity 21.1%; Pred. No. 26;
Matches 30; Conservative 28; Mismatches 67; Indels 17; Gaps 3;
```

Db 297 RLAAARLQNGKQVGSY-----QLPKGMSSHLNGCARTSSSKLWASKSSATVFQNPNG 351
QY 65 KSETTITQSGYNGAD---VGQADNS-----TIELTQNGFRNNATIDQWAKNSD 112
Db 352 IPSSPKNAGYKNSLERNIKQANNSLLHLKLSOTIPKPMNGHSHSIRGSIFSESSTP 411
QY 113 ITVQYGGNNAALVNQTSDDSS 134
Db 412 TTIDEYDNNPSTDDSGDES 433

RESULT 9

US-09-806-709-69
; Sequence 69, Application US/09806709
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: YANG, Yan-Ping
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS INFLUENZAE HIGH
; FILE REFERENCE: 1038-1138 MIS
; CURRENT APPLICATION NUMBER: US/09/806,709
; CURRENT FILING DATE: 2001-04-04
; PRIOR APPLICATION NUMBER: PCT/CA99/00938
; PRIOR FILING DATE: 1999-10-07
; PRIOR APPLICATION NUMBER: 09/167,568
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: 09/206,942
; PRIOR FILING DATE: 1999-12-08
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 69
; LENGTH: 1095
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-806-709-69

Query Match 10.6%; Score 80.5; DB 5; Length 1095;
Best Local Similarity 26.4%; Pred. No. 30;
Matches 39; Conservative 22; Mismatches 60; Indels 27; Gaps 7;
QY 15 SGSALAGVYDQ--LVTRVTH---EMAHASGPDSTLSIYQGSANAALALQS-DARKSET 68
Db 727 TGIILGIESSSGSVTLTATEGALAVSNISG--NTVTANSALTLAGSTIKGTESVT 784
QY 69 TITQSGYNGADVQGGADNSTIEL-----TQNGFRNNATIDQWAKNSDITVQYGGN 121
Db 785 TSSQSG-----DIGGTISGGTVEVKATESLTQNSKIKATGTEANVTSGTIGGTISG 839
QY 122 NAALVNQTSDDSSVMVRQVFGNNATAN 149
Db 840 NTNVNTANAGDLTV-----GNGAEIN 860

RESULT 10

US-09-806-709-67
; Sequence 67, Application US/09806709
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: YANG, Yan-Ping
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS INFLUENZAE HIGH
; FILE REFERENCE: 1038-1138 MIS
; CURRENT APPLICATION NUMBER: US/09/806,709
; CURRENT FILING DATE: 2001-04-04
; PRIOR APPLICATION NUMBER: PCT/CA99/00938
; PRIOR FILING DATE: 1999-10-07
; PRIOR APPLICATION NUMBER: 09/167,568
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: 09/206,942
; PRIOR FILING DATE: 1999-12-08
; NUMBER OF SEQ ID NOS: 95

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 67
; LENGTH: 1536
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-806-709-67

Query Match 10.6%; Score 80.5; DB 5; Length 1536;
Best Local Similarity 26.4%; Pred. No. 46;
Matches 39; Conservative 22; Mismatches 60; Indels 27; Gaps 7;
QY 15 SGSALAGVYDQ--LVTRVTH---EMAHASGPDSTLSIYQGSANAALALQS-DARKSET 68
Db 1168 TGIILGIESSSGSVTLTATEGALAVSNISG--NTVTANSALTLAGSTIKGTESVT 1225
QY 69 TITQSGYNGADVQGGADNSTIEL-----TQNGFRNNATIDQWAKNSDITVQYGGN 121
Db 1226 TSSQSG-----DIGGTISGGTVEVKATESLTQNSKIKATGTEANVTSGTIGGTISG 1280
QY 122 NAALVNQTSDDSSVMVRQVFGNNATAN 149
Db 1281 NTNVNTANAGDLTV-----GNGAEIN 1301

RESULT 11

US-10-170-205E-11455
; Sequence 11455, Application US/10170205E
; GENERAL INFORMATION:
; APPLICANT: ADAMS, Mark
; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN
; FILE REFERENCE: CL001381
; CURRENT APPLICATION NUMBER: US/10/170,205E
; CURRENT FILING DATE: 2002-06-13
; NUMBER OF SEQ ID NOS: 40312
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 11455
; LENGTH: 485
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-170-205E-11455

Query Match 10.6%; Score 80; DB 6; Length 485;
Best Local Similarity 28.3%; Pred. No. 12;
Matches 35; Conservative 14; Mismatches 53; Indels 22; Gaps 5;
QY 32 THEWAHA-----SGPDSLSIYQGSANAALALQS-DARKSETTIT--QSGYNGA 79
Db 276 THSIAHLVPSQTVVQTFSPNPDGTVSLIQGT-GATVATLADASELPTTIVTAQVNTYSAVA 334
QY 80 DVQGGADN-----STIELTQNGFRNNATIDQWAKNSDITVQYGGNNAALVNQTS 131
Db 335 D-GEVEQNWATLOGGENTICTTQASEATQAVASLAEEAAVAASQEMQGGATVTTALNSEAA 393
QY 132 DSSV 135
Db 394 AHAV 397

RESULT 12

US-10-170-205E-11130
; Sequence 11130, Application US/10170205E
; GENERAL INFORMATION:
; APPLICANT: ADAMS, Mark
; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN
; FILE REFERENCE: CL001381
; CURRENT APPLICATION NUMBER: US/10/170,205E
; CURRENT FILING DATE: 2002-06-13
; NUMBER OF SEQ ID NOS: 40312
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 11130
; LENGTH: 503

```
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-170-205E-11130

Query Match      10.6%; Score 80; DB 6; Length 503;
Best Local Similarity 28.2%; Pred. No. 13;
Matches 35; Conservative 14; Mismatches 53; Indels 22; Gaps 5;

QY 32 THEMABA-----SGPSTLSIYQGSANAALALQSDARKSETTIT--QSGYNGA 79
Db 294 THSIAHLVPSQTVVQTFNSPBDGTVSLIQVGT-GATVATLADASELPITVTVAQVNSAVA 352
QY 80 DYVGQADN-----STIELTQNGFRNNATIDQWNAKNSDITVGYGNNAAALVNQTAS 131
Db 353 D-GEVEQWATLQGGEMTIQTQASEATCAVASUAEEAAVAASQEQGATVTMALNSEAA 411
QY 132 DSSV 135
Db 412 AHAV 415

RESULT 13
US-10-170-205E-111456
; Sequence 11456, Application US/10170205E
; GENERAL INFORMATION:
; APPLICANT: ADAMS, Mark
; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN
; FILE REFERENCE: CLO01381
; CURRENT APPLICATION NUMBER: US/10/170,205E
; CURRENT FILING DATE: 2002-06-13
; NUMBER OF SEQ ID NOS: 40312
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 11456
; LENGTH: 503
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-170-205E-11456

Query Match      10.6%; Score 80; DB 6; Length 503;
Best Local Similarity 28.2%; Pred. No. 13;
Matches 35; Conservative 14; Mismatches 53; Indels 22; Gaps 5;

QY 32 THEMABA-----SGPSTLSIYQGSANAALALQSDARKSETTIT--QSGYNGA 79
Db 294 THSIAHLVPSQTVVQTFNSPBDGTVSLIQVGT-GATVATLADASELPITVTVAQVNSAVA 352
QY 80 DYVGQADN-----STIELTQNGFRNNATIDQWNAKNSDITVGYGNNAAALVNQTAS 131
Db 353 D-GEVEQWATLQGGEMTIQTQASEATCAVASUAEEAAVAASQEQGATVTMALNSEAA 411
QY 132 DSSV 135
Db 412 AHAV 415

RESULT 14
US-10-425-115-362763
; Sequence 362763, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 362763
; LENGTH: 546
; TYPE: PRT
```

```
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(546)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_94013C.1.pep
US-10-425-115-362763

Query Match      10.6%; Score 80; DB 6; Length 546;
Best Local Similarity 27.2%; Pred. No. 14;
Matches 34; Conservative 14; Mismatches 43; Indels 34; Gaps 5;

QY 1 MKLLKVAAPAAIIVVSGSALAGVYDQLVTRVVTHEMAHAGSPDSTLS-IYQGSANAALAL 59
Db 90 LMLRLGAVPTLVVSSPRAA-----EATIR--TDHVFAFRPTSTLSDEILYGSDFAFSP 143
QY 60 QSD-----ARKSETTITQSGYNGADYGGQADNSTIELTON 95
Db 144 YGEHWFQARKLVTAHLFTVKRVHSYRRARKEEVRLVVAKVREAAVAGTATDMS---LAMN 200
QY 96 GFRNN 100
Db 201 TFAND 205

RESULT 15
US-60-579-902-7335
; Sequence 7335, Application US/60579902
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Krasomil-Osterfeld, Karina C.
; APPLICANT: Wu, Wei
; TITLE OF INVENTION: Nucleotide and Amino Acid Sequences from Xenorhabdus
; FILE REFERENCE: 38-21 (53444) A
; CURRENT APPLICATION NUMBER: US/60/579,902
; CURRENT FILING DATE: 2004-06-15
; NUMBER OF SEQ ID NOS: 14985
; SEQ ID NO 7335
; LENGTH: 829
; TYPE: PRT
; ORGANISM: Xenorhabdus bovienii
US-60-579-902-7335

Query Match      10.6%; Score 80; DB 7; Length 829;
Best Local Similarity 25.2%; Pred. No. 24;
Matches 36; Conservative 24; Mismatches 53; Indels 30; Gaps 6;

QY 23 YDQLVTRVVTHEMAHAGSPDSTLSIYQGSANAALALQSD-ARKSETTITQSGY-NGAD 80
Db 474 HDEL-----SYQKHKRFQLNLSQNLSDYGSIVYVGYLQDDWGSRTSTRSLNLYGSYNYAD 528
QY 81 VGQGADNSTIELTQNGFRNNATIDQWNAKNSDITVGY-----GGNNA 124
Db 529 I-----NYILNYTONMSSGNSSTADNQLAFNIQVPLDRWLPNSWATISLHNHRCGDNQSA 583
QY 125 LVNQTSADSSVM--VRQVGFNN 145
Db 584 GLSGTALDGNVGLGYLQQRYYNN 606

Search completed: August 2, 2004, 15:29:52
Job time : 18.8 secs
```

This Page Blank (uspto)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:39:53 ; Search time 9.4 Seconds
(without alignments)
1545.204 Million cell updates/sec

Title: US-09-543-407-16

Perfect score: 757

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVGFNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_78.*

1: Piri.*

2: Piri2.*

3: Piri3.*

4: Piri4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	658	86.9	151	2	JC6039
2	658	86.9	151	2	A10635
3	484	63.9	151	2	S70788
4	474.5	62.7	152	2	D90806
5	474.5	62.7	152	2	H85665
6	118.5	15.7	151	2	S70787
7	118.5	15.7	151	2	C90806
8	118.5	15.7	151	2	G85665
9	111.5	14.7	151	2	JC6040
10	111.5	14.7	151	2	AH0635
11	101.5	13.4	590	1	A45621
12	101	13.3	599	2	B42049
13	101	13.3	599	2	A4951
14	99.5	13.1	582	2	F70675
15	99.5	13.1	646	1	S19916
16	98	12.9	145	2	A3143
17	98	12.9	145	2	H38144
18	97.5	12.7	1034	2	JC2143
19	96.5	12.7	602	1	PL0221
20	96	12.7	1651	2	JC1340
21	95.5	12.6	1258	2	J00188
22	94.5	12.5	1322	2	S07053
23	93.5	12.4	1567	2	S11672
24	91.5	12.1	1655	2	E37835
25	89.5	11.8	552	2	D70604
26	89.5	11.8	539	2	C42049
27	88.5	11.7	590	2	E70946
28	87	11.5	329	2	S23247
29	87	11.5	331	2	S21406

outer membrane prote
hypothetical prote
F3F19.21 protein -
probable PPE prote
ice nucleation pro
ice nucleation pro
nucleoskeletal-lik
probable PPE prote
ice nucleation pro
hypothetical prote
hypothetical prote
paracrystalline su
S-layer protein Rs
hemolysin importe
cell wall surface
hypothetical prote
hypothetical prote

30 87 11.5 331 2 S21408
31 87 11.5 455 2 C29349
32 87 11.5 573 2 C86266
33 86.5 11.4 3300 2 D70575
34 85.5 11.3 1210 2 A25547
35 84 11.1 823 2 S14055
36 84 11.1 823 2 E70969
37 83.5 11.0 1200 1 SNPSO
38 82 10.9 528 2 S69589
39 82 10.8 534 2 T32020
40 82 10.8 1026 2 A48995
41 82 10.8 1073 2 C87374
42 82 10.8 1635 2 A10452
43 82 10.8 4776 2 E95206
44 81.5 10.8 760 2 C90739
45 81.5 10.8 760 2 E85589

ALIGNMENTS

RESULT 1

fimbrin protein agfa precursor - Salmonella enteritidis
JC6039
C:Species: Salmonella enteritidis
C:Date: 31-Dec-1996 #sequence.revision 31-Dec-1996 #text_change 08-Oct-1999
C:Accession: JC6039; PC6015; A44898
R:Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Baner, P.A.; Kay, W.W.
J. Bacteriol. 178, 662-667, 1996
A:Title: Salmonella enteritidis agfBAC operon encoding thin, aggregative fimbrinae.
A:Reference number: JC6039; MUID:96146512; PMID:8550497
A:Accession: JC6039
A:Molecule type: DNA
A:Residues: 1-151 <COL>
A:Cross-references: GB:U43280; NID:gl184712; PIDN:AAC43599.1; PID:gl184714
A:Accession: PC6015
A:Molecule type: protein
A:Residues: 21-52 <CO2>
A:Experimental source: strain 27655-3b
A:Note: The authors translated the codon ACG for residue 44 as Ile
R:Collinson, S.K.; Emody, L.; Muller, K.H.; Trust, T.J.; Kay, W.W.
J. Bacteriol. 173, 4773-4781, 1991
A:Title: Purification and characterization of thin, aggregative fimbrinae from Salmonella
A:Reference number: A44898; MUID:91310586; PMID:1677357
A:Contents: 27655
A:Accession: A44898
A>Status: preliminary
A:Molecule type: protein
A:Residues: 21-33 <CO3>
A:Note: sequence extracted from NCBI backbone (NCBIP:45936)
C:Genetics:
A:Gene: agfa
C:Function:
A:Description: major component of thin aggregative fimbrinae
A:Note: fimbrinae bind to fibronectin, plasminogen, tissue plasminogen activator
C:Keywords: fimbrina
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-151/Product: fimbrin protein agfa #status experimental <MAT>

Query Match 86.9%; Score 658; DB 2; Length 151;

Best Local Similarity 90.1%; Pred. No. 7.5e-52;

Matches 136; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

Qy 1 MKLLKVAFAAIVVSGSALAGVYDQLVTRVTHEMAHAGSPDSTLSIYGYGSANAALQ 60

Db 1 MKLLKVAFAAIVVSGSALAGVYDQLVTRVTHEMAHAGSPDSTLSIYGYGSANAALQ 60

Qy 61 SDARKSETTITSGYGNAGADVGGADNSTIETQNGFRNNATIDOWNAKNSDITVGYGG 120

Db 61 SDARKSETTITSGYGNAGADVGGADNSTIETQNGFRNNATIDOWNAKNSDITVGYGG 120

Qy 121 NNAALVNTASDSSVMVRQVGFNNATANQY 151

Db 121 NNAALVNTASDSSVMVRQVGFNNATANQY 151

Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 2

AI0635

major curlin chain precursor [imported] - Salmonella enterica subsp. enterica serovar Typhimurium

C:Species: Salmonella enterica subsp. enterica serovar typhi

A:Note: this species has also been called Salmonella typhi

C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002

C:Accession: AI0635

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, C.; Parkhill, J.; Cronin, A.; Davies, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, M.; Moule, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A:Authors: Parkhill, J.; Cronin, A.; Davies, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, M.; Moule, S.; O'Gaora, P.

A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhimurium

A:Reference number: AB0502; MUID:21534947; PMID:11677608

A:Accession: AI0635

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-151 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD08268.1; PID:g16502315; GSPDB:GN00176

C:Genetics:

A:Gene: STY1181

Query Match 86.9%; Score 658; DB 2; Length 151;

Best Local Similarity 90.1%; Pred. No. 7.5e-52;

Matches 136; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

Qy 1 MKLLKVAFAAIVVSGSALAGVYDQLVTRVVTTHMAHAGSPDSTLSIYQGSANAALALQ 60

Db 1 MKLLKVAFAAIVVSGSALAGVYDQLVTRVVTTHMAHAGSPDSTLSIYQGSANAALALQ 60

Qy 61 SDARKSETTITQSGYNGADVGCGADNSTIETQNGFRNNATIDOWNAKNSDITVGOYGG 120

Db 61 SDARKSETTITQSGYNGADVGCGADNSTIETQNGFRNNATIDOWNAKNSDITVGOYGG 120

Qy 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 3

S70788

curlin protein csgA precursor - Escherichia coli (strain K-12)

N:Alternate names: csgA protein; major curlin protein

C:Species: Escherichia coli

C:Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 01-Mar-2002

C:Accession: S70788; G64846; S31202; S34560; S34559

R:Hammar, M.; Arngqvist, A.; Bian, Z.; Olsen, A.; Normark, S.

Mol. Microbiol. 18, 661-670, 1995

A:Title: Expression of two csg operons is required for production of fibronectin- and curli

A:Reference number: S70783; MUID:96414468; PMID:8817489

A:Accession: S70788

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-151 <HAM>

A:Cross-references: EMBL:X90754; NID:g1147558; PIDN:CAA62282.1; PID:g1147564

A:Experimental source: strain K12, substrain W3110

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, D.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: G64846

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-151 <BLAT>

A:Cross-references: GB:AE000205; GB:U00096; NID:g1787265; PIDN:AAC74126.1; PID:g1787279;

A:Experimental source: strain K-12, substrain MG1655

R:Olsen, A.; Arngqvist, A.; Hammar, M.; Sukupolvi, S.; Normark, S.

Mol. Microbiol. 7, 523-536, 1993

A:Title: The RpoS sigma factor relieves H-NS-mediated transcriptional repression of csgA

A:Reference number: S31202; MUID:93211294; PMID:8459772

A:Accession: S31202

A:Molecule type: DNA

A:Residues: 1-6, 'V', 8-151 <OLS1>

A:Cross-references: EMBL:L04979

A:Accession: S34560

A:Molecule type: protein

A:Residues: 21-42; 44-50 <OLS2>

R:Olsen, A.N.; Arngqvist, A.M.

submitted to the EMBL Data Library, October 1992

A:Reference number: S34559

A:Accession: S34559

A:Molecule type: DNA

A:Residues: 1-133, 'RQRDSGLWV' <OLS3>

A:Cross-references: EMBL:L04979; NID:g290424; PIDN:AAA23616.1; PID:g290425

A:Experimental source: strain K-12, substrain W3110

C:Genetics:

A:Gene: csgA

A:Map position: 23.15

C:Function:

A:Description: major component of wild-type curli; interaction between CsgA and CsgB tri

A:Note: curli are thin, coiled fibers expressed on the surface of Escherichia coli that

and H-kininogen; in the absence of CsgA, CsgB can self-assemble into polymers

F:1-20/Domain: signal sequence #status predicted <SIG>

F:21-151/Product: curlin #status experimental <MAT>

Query Match 63.9%; Score 484; DB 2; Length 151;

Best Local Similarity 66.3%; Pred. No. 2.7e-36;

Matches 101; Conservative 18; Mismatches 32; Indels 0; Gaps 0;

Qy 1 MKLLKVAFAAIVVSGSALAGVYDQLVTRVVTTHMAHAGSPDSTLSIYQGSANAALALQ 60

Db 1 MKLLKVAFAAIVVSGSALAGVYDQLVTRVVTTHMAHAGSPDSTLSIYQGSANAALALQ 60

Qy 61 SDARKSETTITQSGYNGADVGCGADNSTIETQNGFRNNATIDOWNAKNSDITVGOYGG 120

Db 61 TDARNSTLTITQGGNGADVGCGDSSIDLQRFSGNSLTDQNGKNSMTVQFGG 120

Qy 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

Db 121 GNGRAVDQTASNSVNVTVQVGFNNATANQY 151

RESULT 4

D90806

curlin major subunit CsgA [imported] - Escherichia coli (strain O157:H7, substrain RIMD

C:Species: Escherichia coli

C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001

C:Accession: D90806

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.

gawara, N.; Yasunaga, T.; Kuwara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene

A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: D90806

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-152 <HAY>

A:Cross-references: GB:BA000007; PIDN:BA34843.1; PID:g13360880; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RIMD 0509952

C:Genetics:

A:Gene: ECs1420

Query Match 52.7%; Score 474.5; DB 2; Length 152;

Best Local Similarity 55.8%; Pred. No. 2e-35;

Matches 100; Conservative 19; Mismatches 32; Indels 1; Gaps 1;

Qy 1 MKLLKVAFAAIVVSGSALAGVYDQLVTRVVTTHMAHAGSPDSTLSIYQGSANAALAL 59

Db 1 MKLLKVAFAAIVVSGSALAGVYDQLVTRVVTTHMAHAGSPDSTLSIYQGSANAALAL 60

Qy 60 QSDARKSETTITQSGYNGADVGCGADNSTIETQNGFRNNATIDOWNAKNSDITVGOYGG 119

Db 61 QADARNSLDITIQHGGNGADVGGQSDSSIDLQTFGFGNSATLDQWNGKDSHTVTKQFG 120
 QY 120 GNNAAVLNQTASDSSVMVROVGFNNATANQY 151
 Db 121 GNGAAVQQTASNTVNTVQVGFNNATAHQY 152
 RESULT 5
 H55665
 hypothetical protein csgA [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
 C:Species: Escherichia coli
 C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 C:Accession: H55665
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
 Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouisis, K.; Apodaca,
 Nature 409, 529-533, 2001
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A:Reference number: A85480; MUID:21074935; PMID:11206551
 A:Accession: H55665
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-151 <STO>
 A:Cross-references: GB:AE005174; NID:gi2514574; PIDN:AAG55788.1; GSPDB:GN00145; UWGP:Z16
 A:Experimental source: strain O157:H7, substrain EDL933
 C:Genetics:
 A:Gene: csgA
 Query Match 62.7%; Score 474.5; DB 2; Length 152;
 Best Local Similarity 65.8%; Pred. No. 2e-35;
 Matches 100; Conservative 19; Mismatches 33; Indels 1; Gaps 1;
 QY 1 MKLLKVAFAAIVGSGSALAGYDQLVTRVTH-EMAHASGPDSTLSIYQGSANAALAL 59
 Db 1 MKLLKVAFAAIVFGSALAGYVPOYGGGGNGHGGGNNSGFNSSELNIYQVGGNSALAL 60
 QY 60 QSDARKSETTITQSGYNGADVGGADNSTTIELTQNGFRNATIDQWNAKNSDITVGYG 119
 Db 61 QADARNSLDITIQHGGNGADVGGQSDSSIDLQTFGFGNSATLDQWNGKDSHTVTKQFG 120
 QY 120 GNNAAVLNQTASDSSVMVROVGFNNATANQY 151
 Db 121 GNGAAVQQTASNTVNTVQVGFNNATAHQY 152
 RESULT 6
 S70787
 curlin nucleator protein csgB precursor - Escherichia coli (strain K-12)
 N:Alternate names: csgB protein; curlin nucleation component; minor curlin protein
 C:Species: Escherichia coli
 C>Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 01-Mar-2002
 C:Accession: S70787; F64846
 R:Hammar, M.; Arnqvist, A.; Bian, Z.; Olsen, A.; Normark, S.
 Mol. Microbiol. 18, 661-670, 1995
 A:Title: Expression of two csg operons is required for production of fibronectin- and Co
 A:Reference number: S70783; MUID:96414468; PMID:8817489
 A:Accession: S70787
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-151 <HAM>
 A:Cross-references: EMBL:X90754; NID:gi147559; PIDN:CAA62281.1; PID:gl147563
 A:Experimental source: strain K12, substrain W3110
 A>Note: The nucleotide sequence was submitted to the EMBL Data Library, August 1995
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
 A.; Rose, D.J.; Mau, B.; Shaoh, Y.
 Science 277, 1453-1462, 1997
 A:Title: The complete genome sequence of Escherichia coli K-12.
 A:Reference number: A64720; MUID:97426617; PMID:9278503
 A:Accession: F64846
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-151 <BLAT>
 A:Cross-references: GB:AE000205; GB:U000096; NID:gi1787265; PIDN:AAC74125.1; PID:gl1787278;

A:Experimental source: strain K-12, substrain MG1655
 C:Genetics:
 A:Gene: csgB
 A:Map position: 23.15
 C:Function:
 A:Description: minor component of wild-type curli; interaction between CsgA and CsgB tr
 A>Note: curli are thin, coiled fibers expressed on the surface of Escherichia coli that
 and H-kininogen; in the absence of CsgA, CsgB can self-assemble into polymers
 F:1-21/Domain: signal sequence #status predicted <SIG>
 F:22-151/Product: minor curlin chain #status predicted <MAT>
 Query Match 15.7%; Score 118.5; DB 2; Length 151;
 Best Local Similarity 31.8%; Pred. No. 0.0014;
 Matches 37; Conservative 13; Mismatches 62; Indels 5; Gaps 2;
 QY 35 MAHASGPDSTLSIYQGSANAALALQSDARKSETTITQSGYNGADVGGADNSTTIELTQ 94
 Db 18 IAAAGYDLANSEYNF----AVNELSKSFNQAAIIGQAGTNNSAQLRGGSKLLAVVAQ 73
 QY 95 NGFRNNATIDQWNAKNSDITVGYGGNNAALVNQTASDSSVMVROVGFNNATANQY 151
 Db 74 EGSSNRAKIDQTGDYNL-AVIDQAGSANDASISQAGYNTAMTIQKSGNKANITQY 129
 RESULT 7
 C90806
 minor curlin, subunit precursor CsgB [imported] - Escherichia coli (strain O157:H7, subs
 C:Species: Escherichia coli
 C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
 C:Accession: C90806
 R:Hayaishi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G
 Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 8, 11-22, 2001
 A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen
 A:Reference number: A99629; MUID:21156231; PMID:11258796
 A:Accession: C90806
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-151 <HAY>
 A:Cross-references: GB:BA000007; PIDN:BAB34842.1; PID:gi33360879; GSPDB:GN00154
 A:Experimental source: strain O157:H7, substrain RIMD 050952
 C:Genetics:
 A:Gene: ECG1419
 Query Match 15.7%; Score 118.5; DB 2; Length 151;
 Best Local Similarity 31.8%; Pred. No. 0.0014;
 Matches 37; Conservative 13; Mismatches 62; Indels 5; Gaps 2;
 QY 35 MAHASGPDSTLSIYQGSANAALALQSDARKSETTITQSGYNGADVGGADNSTTIELTQ 94
 Db 18 IAAAGYDLANSEYNF----AVNELSKSFNQAAIIGQAGTNNSAQLRGGSKLLAVVAQ 73
 QY 95 NGFRNNATIDQWNAKNSDITVGYGGNNAALVNQTASDSSVMVROVGFNNATANQY 151
 Db 74 EGSSNRAKIDQTGDYNL-AVIDQAGSANDASISQAGYNTAMTIQKSGNKANITQY 129
 RESULT 8
 G85665
 curlin minor chain precursor, CsgA homolog [imported] - Escherichia coli (strain O157:H
 C:Species: Escherichia coli
 C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 C:Accession: G85665
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhe
 Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouisis, K.; Apodaca
 Nature 409, 529-533, 2001
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A:Reference number: A85480; MUID:21074935; PMID:11206551
 A:Accession: G85665
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-151 <STO>
 A:Cross-references: GB:AE005174; NID:gi2514573; PIDN:AAG55787.1; GSPDB:GN00145; UWGP:Z1

A;Experimental source: strain O157:H7, substrain EDL933

C;Genetics:

A;Gene: *csbB*

Query Match 15.7%; Score 118.5; DB 2; Length 151;
Best Local Similarity 31.6%; Pred. No. 0.0014;
Matches 37; Conservative 13; Mismatches 62; Indels 5; Gaps 2;

Qy 35 MAHASGPDTLSIYOGSANAALQSDARKSETTITQSGYNGADVQO-GADNSTIETUQ 94
Db 18 IIAAGYDILANEYNF----AVNELSKSFNOAALIGQGTNNSQLQGGSKLLAVQAQ 73
Qy 95 NGFRNNATIDQNAKNSDITVQYCGNNAALVNQTSVVMVRQVGFNNATANQY 151
Db 74 EGSSNRAKIDQTDYNL-AYIDQAGSANDASISQAGYNTAMIIQKSGNKANITQY 129

RESULT 9

JC6040

fimbriin protein agfB precursor - *Salmonella enteritidis*

C;Species: *Salmonella enteritidis*

C;Date: 31-Dec-1996 #sequence_revision 31-Dec-1996 #text_change 08-Oct-1999

C;Accession: JC6040

R;Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Baner, P.A.; Kay, W.W.

J. Bacteriol. 179, 662-667, 1996

A;Title: *Salmonella enteritidis* agfBAC operon encoding thin, aggregative fimbriae.

A;Reference number: JC6039; MUID:96146512; PMID:8550497

A;Accession: JC6040

A;Molecule type: DNA

A;Residues: 1-151 <COL>

A;Cross-references: GB:U43280; NID:gl184712; PIDN:ARC43598.1; PID:gl184713

A;Experimental source: strain 276755-3b

C;Genetics:

A;Gene: *agfB*

C;Function:

A;Description: minor component of thin aggregative fimbriae

A;Note: fimbriae bind to fibronectin, plasminogen, tissue plasminogen activator

C;Keywords: fimbria

F;1-21/Domain: signal sequence #status predicted <SIG>

F;22-151/Product: fimbriin protein agfB #status predicted <MAT>

Query Match 14.7%; Score 111.5; DB 2; Length 151;
Best Local Similarity 30.5%; Pred. No. 0.0057;
Matches 36; Conservative 17; Mismatches 46; Indels 19; Gaps 5;

Qy 51 GSANAALQSDARKSE-----TTITQSGYNGADVQO-GADNST-----IELTQ 94
Db 14 GAPGIATATNYDLARSEYNFAVNELSKSFNOAALIGQGTNSARVRQEGSKLLSVISQ 73

Qy 95 NGFRNNATIDQNAKNSDIT-VQYCGNNAALVNQTSVVMVRQVGFNNATANQY 151
Db 74 EGGNNRAKVQDQ--AGNYNFAYIEQTGNANDASISQAYGNSAAIIQKSGNKANITQY 129

RESULT 10

AH0635

nucleation component of curlin monomers [imported] - *Salmonella enterica* subsp. *enterica*

C;Species: *Salmonella enterica* subsp. *enterica* serovar Typhi

A;Note: this species has also been called *Salmonella typhi*

C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002

C;Accession: AH0635

R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,

th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,

S.; Moulle, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.

A;Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serov

A;Reference number: AB0502; MUID:21534947; PMID:11677608

A;Accession: AH0635

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-151 <PAR>

A;Cross-references: GB:AL513382; PIDN:CAD08267.1; PID:gl6502314; GSPDB:GN00176

C;Genetics:

A;Gene: *STY1180*

Query Match 14.7%; Score 111.5; DB 2; Length 151;
Best Local Similarity 30.5%; Pred. No. 0.0057;
Matches 36; Conservative 17; Mismatches 46; Indels 19; Gaps 5;

Qy 51 GSANAALQSDARKSE-----TTITQSGYNGADVQO-GADNST-----IELTQ 94
Db 14 GAPGIATATNYDLARSEYNFAVNELSKSFNOAALIGQGTNSARVRQEGSKLLSVISQ 73

Qy 95 NGFRNNATIDQNAKNSDIT-VQYCGNNAALVNQTSVVMVRQVGFNNATANQY 151
Db 74 EGGNNRAKVQDQ--AGNYNFAYIEQTGNANDASISQAYGNSAAIIQKSGNKANITQY 129

RESULT 11

A45621

Leishmanolysin (EC 3.4.24.36) precursor - *Leishmania donovani*

N;Alternate names: surface endopeptidase glycoprotein gp63

C;Species: *Leishmania donovani*

C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 28-Jan-2000

C;Accession: A45621

R;Webb, J.R.; Button, L.L.; McMaster, W.R.

Mol. Biochem. Parasitol. 48, 173-184, 1991

A;Title: Heterogeneity of the genes encoding the major surface glycoprotein of *Leishman*

A;Reference number: A45621; MUID:92107220; PMID:1762629

A;Accession: A45621

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-590 <WEB>

A;Experimental source: LV9

A;Note: sequence extracted from NCBI backbone (NCBIN:74958, NCBI:74959)

C;Function:

A;Description: catalyzes the hydrolysis of peptide bonds between two hydrophobic residu

A;Note: the activated form can activate the proenzyme form

C;Superfamily: leishmanolysin

C;Keywords: blocked carboxyl end; cell adhesion; glycoprotein; hydrolase; lipoprotein; t

F;1-39/Domain: signal sequence #status predicted <SIG>

F;40-87/Domain: activation peptide #status predicted <MAT>

F;88-565/Product: leishmanolysin #status predicted <MAT>

F;566-590/Domain: carboxyl-terminal propeptide #status predicted <CTP>

F;48-251,255,321/Binding site: zinc, catalytic (Cys, His, His, His) (inhibited) #status

F;112-129,178-217,301-373,380-443,393-412,402-477,454-498,503-553,523-546/Disulfide bon

F;251,255,321/Binding site: zinc, catalytic (His) (active) #status predicted

F;252/Active site: Glu #status predicted

F;287/Binding site: carboxylate (Asn) (covalent) #status predicted

F;585/Modified site: GPI-anchor ethanolamine amidated carboxyl end (Asn) (in mature for

Query Match 13.4%; Score 101.5; DB 1; Length 590;
Best Local Similarity 28.8%; Pred. No. 0.24;
Matches 46; Conservative 20; Mismatches 59; Indels 35; Gaps 9;

Qy 8 AFRAIVVGSALAGVYDOLVTRVVTHEMAHASGPDSTLSIYOGSANAALQSDARK-- 65
Db 226 AVGVINIPAAINIASRDQLVTRVVTHEMAHALG-----FVVFPDRARILESINVRHKOF 281

Qy 66 -----SETTITQS--GYNGA-----DVQCGADNSTIELTQNGFRNNATIDQNAK 109
Db 282 DVPVINSSTAVAKAREQYGGTLEYLEMEDQGGAGSAGSHIKM-----RNAQ-DELMAP 334

Qy 110 NSDITVQYCGNNAALVNQTA-----SDSSVMVRQVGFNNNA 146
Db 335 ASD--AGYYSALTWATFDLGLFYQADFSS-KAEEMPWGGRNA 371

RESULT 12

B42049

Leishmanolysin (EC 3.4.24.36) precursor, stationary phase - *Leishmania chagasi*

N;Alternate names: glycoprotein gp63, stationary phase; surface metalloproteinase, stat

C;Species: *Leishmania chagasi*

C;Date: 10-Mar-1994 #sequence_revision 07-Apr-1994 #text_change 22-Jun-1999

C;Accession: B42049

R;Ramamoorthy, R.; Donelson, J.E.; Paetz, K.E.; Maybodi, M.; Roberts, S.C.; Wilson, M.E.
J. Biol. Chem. 267, 1888-1895, 1992
A>Title: Three distinct RNAs for the surface protease gp63 are differentially expressed
A:Reference number: A42049; MUID:92112918; PMID:1370484
A:Accession: B42049
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-599 <RAM>
A:CROSS-references: GB:M80669; NID:gl59324; PIDN:AAA29236.1; PID:gl59325
C:Function:
A:Description: catalyzes the hydrolysis of peptide bonds between two hydrophobic residue
A>Note: the activated form can activate the proenzyme form
C:Superfamily: leishmanolysin
C:Keywords: cell adhesion; glycoprotein; hydrolase; membrane bound; metalloproteinase; z
F:1-39/Domain: signal sequence #status predicted <SIG>
F:40-97/Domain: activation peptide #status predicted <ATP>
F:98-599/Product: leishmanolysin #status predicted <RAM>
F:48,261,265,331/Binding site: zinc, catalytic (Cys, His, His, His) (inhibited) #status
F:122-139,188-227,311-383,390-451,403-422,412-486,463-507,512-562,532-555/Disulfide bond
F:261,265,331/Binding site: zinc, catalytic (His) (active) #status predicted
F:262/Active site: Glu #status predicted
F:297,394/Binding site: carboxylate (Asn) (covalent) #status predicted

Query Match 13.3%; Score 101; DB 2; Length 599;
Best Local Similarity 30.2%; Pred. No. 0.27;
Matches 32; Conservative 12; Mismatches 34; Indels 28; Gaps 3;
RESULT 14
Qy 8 AFAAIVVSGSALAGVYDQVLTTRVVTHEMAHAGSPDSTLSIYQYGSANAALALQSDAR--- 64
Db 236 AVGVNIPAAINASRYDQVLTTRVVTHEMAHALG-----FSVGFEGARILLESISNVHR 288
Qy 65 -----KSETTITQSGYNGADV-----GQADNSTIEL 92
Db 289 KDFDVPVINSSTAVAKAREQYGCOTLEYLEIEDQGGAGSAGSHKM 334
RESULT 14
F70675
Probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 15-Sep-2003
C:Accession: F70675
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon,
Rajandream, M.A.; Rogers, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Connor, R.; Davies, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sultston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: F70675
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-582 <COL>
A:CROSS-references: GB:Z82098; GB:AL123456; NID:g3261664; PIDN:CAB05045.1; PID:g280737;
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: PPE

Query Match 13.1%; Score 99.5; DB 2; Length 582;
Best Local Similarity 23.4%; Pred. No. 0.35; Indels 55; Gaps 7;
Matches 46; Conservative 19; Mismatches 77; Indels 55; Gaps 7;
RESULT 15
Qy 8 AFAAIVVSG-----SALAGVYDQVLTTRVVTHEMAHAGSPDS----- 43
Db 116 AFAQLVMSNWFGNLAFLIAAEGAYEQWAAADVAAMVGVHSGASAAAEQLVFPQALQOL 175
Qy 44 -TLISIYQYGSANAALALQSDARKSETTI--TQSGYGN-----GADVQ 83
Db 176 PNLGIGNIGNANLGGGTGDLNTGNGNIGNTLGSNGRDANLGSNGNIGNSNVGGNVGN 235
Qy 84 ---GADNSTIEL-----TQGFRRNATIDQWAKNSDITVGVGGNAALVNOTASD--- 132
Db 236 GNFSGNGRAGIPGSGNVGNGNLGNSGNTGNSVCGFNTGNNVGTGAGSNTICA 295
Qy 133 SVMVTEQVGFNNATAN 149
Db 296 GNTGSSNWGFGNNGIGN 312
RESULT 15
S19916
leishmanolysin (EC 3.4.24.36) precursor - Leishmania mexicana
N:Alternate names: surface metalloproteinase glycoprotein gp63
C:Species: Leishmania mexicana
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
R;Medina-Acosta, E.; Kress, R.E.; Russell, D.
C:Accession: S19916; A48564
submitted to the EMBL Data Library, February 1992
A:Description: Structurally distinct genes for the surface protease (gp63) of Leishmani
A:Reference number: S19916

R;Ramamoorthy, R.; Donelson, J.E.; Paetz, K.E.; Maybodi, M.; Roberts, S.C.; Wilson, M.E.
J. Biol. Chem. 267, 1888-1895, 1992
A>Title: Three distinct RNAs for the surface protease gp63 are differentially expressed
A:Reference number: A42049; MUID:92112918; PMID:1370484
A:Accession: B42049
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-599 <RAM>
A:CROSS-references: GB:M80669; NID:gl59324; PIDN:AAA29236.1; PID:gl59325
C:Function:
A:Description: catalyzes the hydrolysis of peptide bonds between two hydrophobic residue
A>Note: the activated form can activate the proenzyme form
C:Superfamily: leishmanolysin
C:Keywords: cell adhesion; glycoprotein; hydrolase; membrane bound; metalloproteinase; z
F:1-39/Domain: signal sequence #status predicted <SIG>
F:40-97/Domain: activation peptide #status predicted <ATP>
F:98-599/Product: leishmanolysin #status predicted <RAM>
F:48,261,265,331/Binding site: zinc, catalytic (Cys, His, His, His) (inhibited) #status
F:122-139,188-227,311-383,390-451,403-422,412-486,463-507,512-562,532-555/Disulfide bond
F:261,265,331/Binding site: zinc, catalytic (His) (active) #status predicted
F:262/Active site: Glu #status predicted
F:297,394/Binding site: carboxylate (Asn) (covalent) #status predicted

Query Match 13.3%; Score 101; DB 2; Length 599;
Best Local Similarity 30.2%; Pred. No. 0.27;
Matches 32; Conservative 12; Mismatches 34; Indels 28; Gaps 3;
RESULT 14
Qy 8 AFAAIVVSGSALAGVYDQVLTTRVVTHEMAHAGSPDSTLSIYQYGSANAALALQSDAR--- 64
Db 236 AVGVNIPAAINASRYDQVLTTRVVTHEMAHALG-----FSVGFEGARILLESISNVHR 288
Qy 65 -----KSETTITQSGYNGADV-----GQADNSTIEL 92
Db 289 KDFDVPVINSSTAVAKAREQYGCOTLEYLEIEDQGGAGSAGSHKM 334
RESULT 13
A44951
Leishmanolysin (EC 3.4.24.36) precursor, log phase - Leishmania chagasi
N:Alternate names: glycoprotein gp63; surface metalloproteinase, log phase
C:Species: Leishmania chagasi
C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 31-Jan-2000
C:Accession: A44951; A42049
R;Miller, R.A.; Reed, S.G.; Parsons, M.
Mol. Biochem. Parasitol. 39, 267-274, 1990
A>Title: Leishmania gp63 molecule implicated in cellular adhesion lacks an Arg-Gly-Asp s
A:Reference number: A44951; MUID:90205976; PMID:2320059
A:Accession: A44951
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-599 <ML>
A:CROSS-references: GB:M28527; NID:gl59322; PIDN:AAA29235.1; PID:gl59323
R;Ramamoorthy, R.; Donelson, J.E.; Paetz, K.E.; Maybodi, M.; Roberts, S.C.; Wilson, M.E.
J. Biol. Chem. 267, 1888-1895, 1992
A>Title: Three distinct RNAs for the surface protease gp63 are differentially expressed
A:Reference number: A42049; MUID:92112918; PMID:1370484
A:Accession: A42049
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-599 <RAM>
A:CROSS-references: GB:M80672; NID:gl59328; PIDN:AAA29238.1; PID:gl59329
A>Note: sequence extracted from NCBI backbone (NCBI:76040, NCBI:76041)
A>Note: the source is designated as Leishmania donovani chagasi
C:Function:
A:Description: catalyzes the hydrolysis of peptide bonds between two hydrophobic residue
A>Note: the activated form can activate the proenzyme form
C:Superfamily: leishmanolysin
C:Keywords: blocked carboxyl end; cell adhesion; glycoprotein; hydrolase; lipoprotein; m
F:1-39/Domain: signal sequence #status predicted <SIG>
F:40-97/Domain: activation peptide #status predicted <ATP>
F:98-574/Product: leishmanolysin #status predicted <RAM>
F:48,261,265,331/Binding site: zinc, catalytic (Cys, His, His, His) (inhibited) #status

A:Accession: S19916
A:Molecule type: mRNA
A:Residues: 1-646 <MED>
A:Cross-references: EMBL:X64394; NID:99559; PIDN:CAA45733.1; PID:99560
R:Medina-Acosta, E.; Kress, R.E.; Russell, D.G.
Mol. Biochem. Parasitol. 57, 31-45, 1993
A:Title: Structurally distinct genes for the surface protease of Leishmania mexicana ara
A:Reference number: A48564; MUID:93149206; PMID:8426614
A:Accession: A48564
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-176, Q' 178-646 <ME2>
A:Cross-references: EMBL:X64394; NID:99559
A:Note: sequence extracted from NCBI backbone (NCBIP:123747)
C:Genetics:
A:Gene: gp63-C1
A:Map position: 700kb chromosomal band
C:Function:
A:Description: catalyzes the hydrolysis of peptide bonds between two hydropho
A:Note: the activated form can activate the proenzyme form
C:Superfamily: leishmanolysin
C:Keywords: cell adhesion; glycoprotein; hydrolase; membrane bound; metalloproteinase; 2
F:1-39/Domain: signal sequence #status predicted <SIG>
F:40-102/Domain: activation peptide #status predicted <ATP>
F:103-646/Product: leishmanolysin #status predicted <MAT>
F:604-620/Domain: transmembrane #status predicted <TM1>
F:48,466,270,336/Binding site: zinc, catalytic (Cys, His, His, His) (inhibited) #status
F:86,297,399,409,433,445,466,501/Binding site: carboxylate (Asn) (covalent) #status pre
F:127-144,193-232,316-388,395-458,408-427,417-492,469-513,518-568,538-561/disulfide bond
F:266,270,336/Binding site: zinc, catalytic (His) (active) #status predicted
F:267/Active site: Glu #status predicted

Query Match 13.1%; Score 99.5; DB 1; Length 646;
Best Local Similarity 30.8%; Pred. No. 0.4;
Matches 40; Conservative 15; Mismatches 66; Indels 9; Gaps 4;
QY 8 AFAAIVVGSALAGVVDQLVTRVVTHEMAHASGPDSTLSIYQYGSANAALALQSDARK-- 65
Db 241 AVGVINIPAAIASRYDQLVTRVVTHEMAHAFVSGTF---FGAVGIVQVEVPHLRKDF 296
QY 66 SETTITQSGYNGADVGGADN-STIELTQNGFRNNA--TIDQWAKNSDITVGQYGGNN 122
Db 297 NVSVITTSSTVAKAREQYGCNSLEYLEIEDQGGAGSAGSHIXVRNAKDELMAPAASAGYY 356
QY 123 AALVNOTASD 132
Db 357 TALTMVAFQD 366

Search completed: August 2, 2004, 14:56:22
Job time : 10.4 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:36:12 ; Search time 5.3 seconds

(without alignments)
1483:508 Million cell updates/sec

Title: US-09-543-407-16

Perfect score: 757

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRFQVGGNNATANY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	658	86.9	151	1 CSGA_SALTY	P55225 salmonella
2	484	63.9	151	1 CSGA_ECOLI	P38307 escherichia
3	474.5	62.7	152	1 CSGA_ECO57	Q30u24 escherichia
4	118.5	15.7	151	1 CSGB_ECOLI	P39828 escherichia
5	111.5	14.7	151	1 CSGB_SALTY	Q8z7m3 salmonella
6	111.5	14.7	151	1 CSGB_SALTY	P55226 salmonella
7	101.5	13.4	590	1 GP63_LEIDO	P32223 leishmania
8	101	13.3	599	1 GP63_LEICH	P15706 leishmania
9	99.5	13.1	646	1 GP63_LEIME	P43150 leishmania
10	99	13.1	1656	1 OMPB_RICUA	Q06653 r outer mem
11	97.5	12.9	1034	1 ICEN_PANAN	Q47879 pantoaea ana
12	96.5	12.7	602	1 GP63_LEIMA	P08148 leishmania
13	95.5	12.6	1258	1 ICEN_ERWHE	P16239 erwinia her
14	94.5	12.5	1322	1 ICEA_PANAN	P20469 pantoaea ana
15	93.5	12.4	1567	1 ICEN_XANCT	P18127 xanthomonas
16	93	12.3	310	1 H8LO_STRMU	Q8cw23 streptococc
17	91.5	12.1	1655	1 OMPB_RICCN	Q9kka3 r outer mem
18	87	11.5	331	1 OMB1_NEIMB	P30689 reisseria m
19	87	11.5	331	1 OMB3_NEIMB	P30689 reisseria m
20	85.5	11.3	1210	1 ICEN_PSEPL	P09815 pseudomonas
21	84.5	11.2	1196	1 ICEN_PSEBX	P33479 pseudomonas
22	84	11.1	823	1 NSP1_YEAST	P14907 saccharomyc
23	83.5	11.0	1148	1 ICEK_PSEBX	Q30611 pseudomonas
24	83.5	11.0	1200	1 ICEN_PSEBX	P06620 pseudomonas
25	82.5	10.9	504	1 NRF1_BRARE	Q90x44 brachydanio
26	82	10.8	955	1 FRU_PROMO	Q8in81 drosophila
27	82	10.8	1025	1 SLAP_CAUCR	P35828 caulobacter
28	81.5	10.8	422	1 HEAD_BPAPS	Q9t1s4 bacterioph
29	81.5	10.8	760	1 YBIL_ECOLI	P25780 escherichia
30	81.5	10.8	1158	1 RL14_HUMAN	P48552 homo sapien
31	81	10.7	856	1 AUE7_AZOVI	Q3zif9 azotobacter
32	80.5	10.6	720	1 GTAC_BREDI	Q915d6 brevundimon
33	80	10.6	503	1 NRF1_HUMAN	Q16656 homo sapien

34 80 10.6 503 1 NRF1_MOUSE
35 80 10.6 621 1 GP63_LEIGU
36 79.5 10.5 311 1 7B4C_PSESP
37 79.5 10.5 1300 1 120K_RICRI
38 79.5 10.5 1654 1 OMPB_RICRI
39 79 10.4 907 1 GLRI_MOUSE
40 79 10.4 907 1 GLRI_RAT
41 78 10.3 582 1 Y281_MYCPN
42 78 10.3 1643 1 OMPB_RICPR
43 77 10.2 331 1 OMB_NEIMB
44 76.5 10.1 388 1 ARGJ_CORGL
45 76 10.0 303 1 PANE_HALNI

ALIGNMENTS

RESULT 1
CSGA_SALTY
ID CSGA_SALTY STANDARD; PRT; 151 AA.
AC P55225;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Major curlin subunit precursor (Fimbrin SEF17).
GN CSGA OR AGFA OR STM1144 OR STY1181 OR T1776.
OS Salmonella typhimurium.
OS Salmonella typhi, and
OS Salmonella enteritidis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
ON NCBI_TaxID=602, 601, 592;
RX [1]
RP SEQUENCE FROM N.A.
RC SPECIES=Salmonella typhimurium; STRAIN=SR-11;
RX MEDLINE=98117058; PubMed=9457880;
RA Romling U., Bian Z., Hammar M., Sierralta W.D., Normark S.;
RT "Curli fibers are highly conserved between Salmonella typhimurium and
RT Escherichia coli with respect to operon structure and regulation.";
RL J. Bacteriol. 180:722-731(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=Salmonella typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=Salmonella typhi; STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrall B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
RN [4]
RP SEQUENCE FROM N.A.
RC SPECIES=Salmonella typhi; STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;

RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
 RT and CT18.";
 RL J. Bacteriol. 185:2330-2337(2003).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S.enteritidis; STRAIN=27655-3B;
 RX MEDLINE=96146312; PubMed=8550457;
 RA Collinson S.K., Clouthier S.C., Doran J.L., Baner P.A., Kay W.W.;
 RT "Salmonella enteritidis agfBAC operon encoding thin, aggregative
 RT fimbriae.";
 RL J. Bacteriol. 178:662-667(1996).
 RN [6]
 RP SEQUENCE OF 21-151 FROM N.A.
 RC SPECIES=S.enteritidis; STRAIN=27655-3B;
 RX MEDLINE=94013373; PubMed=8104955;
 RA Doran J.L., Collinson S.K., Burian J., Santos G., Todd E.C.D.,
 RA Munro C.K., Kay C.M., Baner P.A., Peterkin P.I., Kay W.W.;
 RT "DNA-based diagnostic tests for Salmonella species targeting agfA,
 RT the structural gene for thin, aggregative fimbriae.";
 RL J. Clin. Microbiol. 31:2263-2273(1993).
 RN [7]
 RP SEQUENCE OF 21-33.
 RC SPECIES=S.enteritidis; STRAIN=27655-3B;
 RX MEDLINE=91310586; PubMed=1677357;
 RA Collinson S.K., Emsedy L., Mueller K.-M., Trust T.J., Kay W.W.;
 RT "Purification and characterization of thin, aggregative fimbriae from
 RT Salmonella enteritidis.";
 RL J. Bacteriol. 173:4773-4781(1991).
 CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 CC FIBRONECTIN.
 CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC -----
 CC EMBL; AJ002301; CAA05317.1; -;
 CC EMBL; AF008749; AAL20074.1; -;
 CC EMBL; AL627269; CAD08268.1; -;
 CC EMBL; AE016840; AAC09399.1; -;
 CC EMBL; U43280; AAC43599.1; -;
 CC EMBL; J06039; J06039.
 CC StyGene; SGI0608; csGA.
 CC Fimbrin; Signal; Complete proteome.
 CC SIGNAL 1 20
 CC CHAIN 21 151
 CC CONFLICT 134 151 SVVVRQVGFNNATANYQ -> DSVTQVAS (IN
 CC REF. 6).
 CC SEQUENCE 151 AA; 15305 MW; B7DAC0D16B621359 CRC64;
 CC
 CC Query Match 86.9%; Score 658; DB 1; Length 151;
 CC Best Local Similarity 90.1%; Pred. No. 4.9e-50;
 CC Matches 136; Conservative 3; Mismatches 12; Indels 0; Gaps 0;
 CC
 CC QY 1 MKLLKVAAPAAIVVSGSALAGVYDQVTRVWVTHMAHAGSPDSTLSIYQGSANAALAQ 60
 CC DB 1 MKLLKVAAPAAIVVSGSALAGVYDQVTRVWVTHMAHAGSPDSTLSIYQGSANAALAQ 60
 CC
 CC 61 SPARKSETTITQSGYNGADVGQADNSTIELTQNGFRNATIDQWAKNSDITVGYGG 120
 CC QY 61 SPARKSETTITQSGYNGADVGQADNSTIELTQNGFRNATIDQWAKNSDITVGYGG 120
 CC DB 61 SPARKSETTITQSGYNGADVGQADNSTIELTQNGFRNATIDQWAKNSDITVGYGG 120
 CC
 CC 121 NNAALVNOTASDSSVMVRQVGFNNATANYQ 151
 CC QY 121 NNAALVNOTASDSSVMVRQVGFNNATANYQ 151
 CC DB 121 NNAALVNOTASDSSVMVRQVGFNNATANYQ 151

RESULT 2
 CSGA_ECOLI STANDARD; PRT; 151 AA.
 ID CSGA_ECOLI
 AC P28307;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Major curlin subunit precursor.
 GN CSGA OR B1042.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / W3110;
 RX MEDLINE=93211294; PubMed=8459772;
 RA Olsen A., Arngvist A.;
 RT "The RpoS sigma factor relieves H-NS-mediated transcriptional
 RT repression of csGA, the subunit gene of fibronectin-binding curli in
 RT Escherichia coli.";
 RL Mol. Microbiol. 7:523-536(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MC4100;
 RX MEDLINE=96414468; PubMed=8817489;
 RA Hammar M., Arngvist A., Bian Z., Olsen A., Normark S.;
 RT "Expression of two csG operons is required for production of
 RT fibronectin- and congo red-binding curli polymers in Escherichia coli
 RT K-12.";
 RL Mol. Microbiol. 18:661-670(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shaq Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97061202; PubMed=8905232;
 RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
 RA Sano M., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horiuchi T.;
 RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
 RT corresponding to the 12.7-28.0 min region on the linkage map.";
 RL DNA Res. 3:137-155(1996).
 RN [5]
 RP SEQUENCE OF 21-40.
 RC STRAIN=K12 / YMEI;
 RX MEDLINE=93033873; PubMed=1357528;
 RA Arngvist A., Olsen A., Pfeifer J., Russell D.G., Normark S.;
 RT "The Crl protein activates cryptic genes for curli formation and
 RT fibronectin binding in Escherichia coli HB101.";
 RL Mol. Microbiol. 6:2443-2452(1992).
 RN [6]
 RP SEQUENCE OF 21-31.
 RC MEDLINE=91310586; PubMed=1677357;
 RX Collinson S.K., Emsedy L., Trust T.J., Kay W.W.;
 RT "Purification and characterization of thin, aggregative fimbriae from
 RT Salmonella enteritidis.";
 RL J. Bacteriol. 173:4773-4781(1991).
 CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 CC FIBRONECTIN.
 CC


```

CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L04979; AAA23616.1; -
DR EMBL; X90754; CAA62282.1; -
DR EMBL; AE000205; AAC74126.1; -
DR EMBL; D90741; BAA35832.1; -
DR EMBL; D90742; BAA35840.1; -
DR PIR; S70788; S70788.
DR EcoGene; E311489; csGA.
KW Fimbria; Signal; Complete proteome.
FT SIGNAL 1 20
FT CHAIN 21 151 MAJOR CURLIN SUBUNIT.
FT CONFLICT 7 7 A -> E (IN REF. 1). CRC64;
SQ SEQUENCE 151 AA; 15049 MW; C003470D208D395F CRC64;

Query Match 63.9%; Score 484; DB 1; Length 151;
Best Local Similarity 66.9%; Pred. No. 4.8e-35;
Matches 101; Conservative 18; Mismatches 32; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVYDQVLTTRVTH-EMAHASGPDSTLSIYQGSANAALALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVYDQVLTTRVTH-EMAHASGPDSTLSIYQGSANAALALQ 60
QY 61 SDARKSETTITQSGYGCNADYVQGGADNSTIELTQNGFRNNATIDOWNAKNSDITVGVGG 120
DB 61 TDARNSDLTITQHGCGNADYVQGGSDSSIDLTRQGFNSATLDQNGKXSEMTVKQFGG 120

QY 121 NNAALVNQTSADSSVNVVQVGFNNATANQY 151
DB 121 GNGAAVDQTASNSVNVTVQVGFNNATAHQY 151

RESULT 3
CSGA_ECO57 STANDARD; PRT; 152 AA.
AC Q93U24;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Major curlin subunit precursor.
GN CSGA OR Z1676 OR ECS1420.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / ATCC 43895;
RX MEDLINE=21218556; PubMed=11319125;
RA Uhlich G.A., Keen J.E., Elder R.O.;
RT "Mutations in the csgD promoter associated with variations in curli
RT expression in certain strains of Escherichia coli O157:H7.";
RL Appl. Environ. Microbiol. 67:2367-2370(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postfai G., Hackett J., Link S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).

```

```

[3] SEQUENCE FROM N.A.
RP STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayaishi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasaki C., Sasawara N., Yasunaga T.,
RA Kihara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN.
CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF275733; AA53212.1; -
DR EMBL; AE005315; AG55788.1; -
DR EMBL; AP002554; BAB34843.1; -
DR PIR; D90806; D90806.
DR PIR; H85665; H85665.
KW Fimbria; Signal; Complete proteome.
FT SIGNAL 1 20 BY SIMILARITY.
FT CHAIN 21 152 MAJOR CURLIN SUBUNIT.
FT CONFLICT 152 AA; 15099 MW; EE2D2D94DDE91243 CRC64;
SQ SEQUENCE 152 AA; 15099 MW; EE2D2D94DDE91243 CRC64;

Query Match 62.7%; Score 474.5; DB 1; Length 152;
Best Local Similarity 65.8%; Pred. No. 3.2e-34;
Matches 100; Conservative 19; Mismatches 32; Indels 1; Gaps 1;

QY 1 MKLLKVAFAAIVVSGSALAGVYDQVLTTRVTH-EMAHASGPDSTLSIYQGSANAALAL 59
DB 1 MKLLKVAFAAIVVSGSALAGVYDQVLTTRVTH-EMAHASGPDSTLSIYQGSANAALAL 60
QY 60 QSDARKSETTITQSGYGCNADYVQGGADNSTIELTQNGFRNNATIDOWNAKNSDITVGVGG 119
DB 61 QADARNSDLTITQHGCGNADYVQGGSDSSIDLTRQGFNSATLDQNGKDSHMTVKQFG 120

QY 120 GNNAAVNQTSADSSVNVVQVGFNNATANQY 151
DB 121 GNGAAVDQTASNSVNVTVQVGFNNATAHQY 152

RESULT 4
CSGB_ECOLI STANDARD; PRT; 151 AA.
ID CSGB_ECOLI
AC P39828;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Minor curlin subunit precursor.
GN CSGB OR B1041 OR Z1675 OR ECS1419.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MC4100;
RX MEDLINE=96414468; PubMed=8817489;
RA Hammar M., Arngvist A., Bian Z., Olsen A., Normark S.;
RT "Expression of two csg operons is required for production of
RT fibronectin- and congo red-binding curli polymers in Escherichia coli

```

RT K-12.";
 RL Mol. Microbiol. 18:661-670(1995).
 RP [2]
 RN
 RC SEQUENCE FROM N.A.
 RX STRAIN=X12 / MG1655;
 RA MEDLINE=97426617; PubMed=9278503;
 RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of *Escherichia coli* K-12.";
 RL Science 277:1453-1474(1997).
 RN [3]
 RN SEQUENCE FROM N.A.
 RP STRAIN=X12;
 RX MEDLINE=97061202; PubMed=8905232;
 RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
 RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horiiuchi I.;
 RT "A 718-Kb DNA sequence of the *Escherichia coli* K-12 genome
 RT corresponding to the 12.7-28.0 min region on the linkage map.";
 RL DNA Res. 3:137-155(1996).
 RN [4]
 RN SEQUENCE FROM N.A.
 RP STRAIN=0157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Rosfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Godbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamocousis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7.";
 RL Nature 409:529-533(2001).
 RN [5]
 RN SEQUENCE FROM N.A.
 RP STRAIN=0157:H7 / RIMD 050952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogawara N., Yasunaga T.,
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohemorrhagic *Escherichia coli*
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22(2001).
 RN [6]
 RN SEQUENCE OF 1-21 FROM N.A.
 RP STRAIN=X12;
 RX MEDLINE=95157246; PubMed=7854117;
 RA Arngvist A., Olsen A., Normark S.;
 RT "Sigma S-dependent growth-phase induction of the *csqBA* promoter in
 RT *Escherichia coli* can be achieved in vivo by sigma 70 in the absence
 RT of the nucleoid-associated protein H-NS.";
 RL Mol. Microbiol. 13:1021-1032(1994).
 CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
 CC CURLIN MONOMERS.
 CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X90754; CAA62281.1; -.

DR EMBL; AE000205; AAC74125.1; -.
 DR EMBL; D90741; BAA35831.1; -.
 DR EMBL; AE005315; AAG55787.1; -.
 DR EMBL; AP002554; BAB34842.1; -.
 DR PIR; C90806; C90806.
 DR PIR; G85665; G85665.
 DR PIR; S70787; S70787.
 DR EcoGene; EG12621; csqB.
 KW Fimbria; Signal; Complete proteome.
 FT SIGNAL 1 21 POTENTIAL.
 FT CDS 22 151 MINOR CURLIN SUBUNIT.
 SQ SEQUENCE 151 AA; 15682 MW; B18D266B964014B8 CRC64;
 Query Match 15.7%; Score 118.5; DB 1; Length 151;
 Best Local Similarity 31.6%; Pred. No. 0.0015;
 Matches 37; Conservative 13; Mismatches 62; Indels 5; Gaps 2;
 Qy 35 MAHAGSDPSTLSYQYGSAAALQSDARKSTTTTQSCYGNAGDVGQADNSTIELTQ 94
 Db 18 IAAAGYDLANSEYNF----AVNELSKSFNQAAIIQCAQTNNNAQLRQGSKLLAVVAQ 73
 Qy 95 NGFENNATIDQNAKNSDITVGOYGNNAALVNOTASDSSVMVROVYGFGNNTATNOY 151
 Db 74 EGSNRKAKIDQTGDYNL-AVIDAGSANDASISQAGYGNVTAMIIKSGNKNATIQY 129
 RESULT 5
 CSGB_SALTI STANDARD; PRT; 151 AA.
 AC Q827M5; 2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Minor curlin subunit precursor.
 GN CSGB OR STY1180 OR T1777.
 OS Salmonella typhi.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 OX NCBI_TaxID=601;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=CT18;
 RC MEDLINE=21534947; PubMed=11677608;
 RX Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
 RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
 RA Baker S., Basham D., Brooks K., Chillingworth I., Connor P.,
 RA Cronin A., Davies P., Davies R.M., Dowd L., White N., Farrar J.,
 RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
 RA Krogg A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
 RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
 RA Whitehead S., Barrall B.G.;
 RT "Complete genome sequence of a multiple drug resistant *Salmonella*
 RT enterica serovar typhi CT18.";
 RL Nature 413:848-852(2001).
 RN [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN=Ty2 / ATCC 700931;
 RC MEDLINE=22531367; PubMed=12644504;
 RX Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
 RA Burland V., Kodyani V., Schwartz D.C., Blattner F.R.;
 RT "Comparative genomics of *Salmonella enterica* serovar typhi strains Ty2
 RT and CT18.";
 RL J. Bacteriol. 185:2330-2337(2003).
 CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
 CC CURLIN MONOMERS.
 CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X90754; CAA62281.1; -.


```

DR MEROPS; M08.001; -
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR InterPro; IPR001577; Peptidase M8.
DR Pfam; PF01457; Peptidase M8; 1.
DR PRINTS; PR00782; LSHMANOLYSIN.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydrolase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;
KW Zymogen; Signal; Cell adhesion; GPI-anchor; Lipoprotein.
FT SIGNAL 1 39 POTENTIAL.
FT PROPEP 40 87 ACTIVATION PEPTIDE.
FT CHAIN 88 565 LEISHMANOLYSIN.
FT PROPEP 566 590 REMOVED IN MATURE FORM (BY SIMILARITY).
FT METAL 251 221 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 252 222 BY SIMILARITY.
FT METAL 255 225 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 321 321 ZINC (CATALYTIC) (BY SIMILARITY).
FT DISULFID 112 129 BY SIMILARITY.
FT DISULFID 178 217 BY SIMILARITY.
FT DISULFID 301 373 BY SIMILARITY.
FT DISULFID 380 443 BY SIMILARITY.
FT DISULFID 393 412 BY SIMILARITY.
FT DISULFID 402 477 BY SIMILARITY.
FT DISULFID 454 498 BY SIMILARITY.
FT DISULFID 503 553 BY SIMILARITY.
FT DISULFID 523 546 BY SIMILARITY.
FT CARBOHYD 287 287 N-LINKED (GLCNAC...) (POTENTIAL).
FT LIPID 565 565 GPI-anchor amidated asparagine (By similarity).
SQ SEQUENCE 590 AA; 62950 MW; 0FB315D299659F58 CRC64;

Query Match 13.4%; Score 101.5; DB 1; Length 590;
Best Local Similarity 28.8%; Pred. No. 0.2;
Matches 46; Conservative 20; Mismatches 59; Indels 35; Gaps 9;

QY 8 AFAIVVSGSALAGVYDQVLVTRVVTTHMAHAGSPDSTLSIYQYGSANALALQSDARK-- 65
DB 226 AVGVINIPAANTASRYDQVLVTRVVTTHMAHALG-----FSVFFRDARILESISNVRHKDF 281
QY 66 -----SETTITQS--GYNGA-----DVGQGDNSTIELTQGFNRNATTIDOMNAK 109
DB 282 DVPVNSTAVAKAREQYCGTGLEVMEDQGGAGSAGSHIKM-----RNAQ-DELNAP 334
QY 110 NSDITVQYGGNNAALVNQTA---SDSSVMVRQVGFQNNNA 146
DB 335 ASD--AGVYSALTMAIFQDLGFGYQADFS--KAEEMPWGRNA 371

RESULT 8
GP63-LEICH STANDARD; PRT; 599 AA.
ID GP63 LEICH STANDARD; PRT; 599 AA.
AC F15706;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Leishmanolysin precursor (EC 3.4.24.36) (Cell surface protease)
DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface
DE endopeptidase).
GN GP63.
OS Leishmania chagasi.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=44271;
RN [1]_TaxID=44271;
RP SEQUENCE FROM N.A.
RX MEDLINE=90205976; PubMed=2320059;
RA Miller R.A., Reed S.G., Parsons M.;
RT "Leishmania gp63 molecule implicated in cellular adhesion lacks an
RT Arg-Gly-Asp sequence";
RL Mol. Biochem. Parasitol. 39:267-274 (1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92112918; PubMed=1370484;
RA Ramamoorthy R., Donelson J.E., Paetz K.E., Maybodi M., Roberts S.C.,
RA Wilson M.E.;

```

```

RT RT
RT RT
RT RT
RL J. Biol. Chem. 267:1898-1895(1992).
CC -I- FUNCTION: Has an integral role during the infection of macrophages
CC in the mammalian host.
CC -I- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and
CC P1' and basic residues at P2 and P3'. A model nonapeptide is
CC cleaved at -Ala-Tyr-Lys-Lys-.
CC -I- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC -I- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -I- SIMILARITY: Belongs to peptidase family M8.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to licenses@isb-sib.ch).
CC -----
CC EMBL; M80672; AAA29238.1; -
CC EMBL; M28527; AAA29235.1; -
CC PIR; A44951; A44951.
CC HSP; P08148; 1LML.
CC MEROPS; M08.001; -
CC InterPro; IPR006025; Pept_M_Zn_BS.
CC InterPro; IPR001577; Peptidase_M8.
CC Pfam; PF01457; Peptidase_M8; 1.
CC PRINTS; PR00782; LSHMANOLYSIN.
CC PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydrolase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;
KW Zymogen; Signal; Cell adhesion; GPI-anchor; Lipoprotein.
FT SIGNAL 1 39 POTENTIAL.
FT PROPEP 40 97 ACTIVATION PEPTIDE.
FT CHAIN 98 574 LEISHMANOLYSIN.
FT PROPEP 575 599 REMOVED IN MATURE FORM (BY SIMILARITY).
FT METAL 261 261 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 262 262 BY SIMILARITY.
FT METAL 265 265 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 331 331 ZINC (CATALYTIC) (BY SIMILARITY).
FT DISULFID 122 139 BY SIMILARITY.
FT DISULFID 188 227 BY SIMILARITY.
FT DISULFID 311 383 BY SIMILARITY.
FT DISULFID 390 452 BY SIMILARITY.
FT DISULFID 403 422 BY SIMILARITY.
FT DISULFID 412 486 BY SIMILARITY.
FT DISULFID 463 507 BY SIMILARITY.
FT DISULFID 512 562 BY SIMILARITY.
FT DISULFID 532 555 BY SIMILARITY.
FT CARBOHYD 297 297 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 394 394 N-LINKED (GLCNAC...) (POTENTIAL).
FT LIPID 574 574 GPI-anchor amidated asparagine (By similarity).
SQ SEQUENCE 599 AA; 63848 MW; 746730AE8E2A2E7C CRC64;

Query Match 13.3%; Score 101; DB 1; Length 599;
Best Local Similarity 30.2%; Pred. No. 0.23;
Matches 32; Conservative 12; Mismatches 34; Indels 28; Gaps 3;

QY 8 AFAIVVSGSALAGVYDQVLVTRVVTTHMAHAGSPDSTLSIYQYGSANALALQSDAR--- 64
DB 236 AVGVINIPAANTASRYDQVLVTRVVTTHMAHALG-----FSVFFRDARILESISNVRH 288
QY 65 -----KSETTITQSGVNGADV-----CQGDNSTIEL 92
DB 289 KDFDVPVNSTAVAKAREQYCGTGLEVMEDQGGAGSAGSHIKM 334

RESULT 9
GP63-LEIME STANDARD; PRT; 646 AA.
ID GP63-LEIME STANDARD; PRT; 646 AA.
AC P43150;

```



```

QY 6 VAAFAIVVSGSALGVYDQLVTRVTHVMAHASGPDSTLSIYQXSANAALALQSDARK 65
DQ 509 VLAAGAITLDGSA-----TITDIGNGG-----GAALQSITLANDATK 547
QY 66 SETTITQSG-----YNGADVGGADNSTIETQNGFRNNATID----- 104
DQ 548 ---TLTGGANITSANGGTINFGANGGKIKLST--QNNIVVDCDLAATDGTGVVDASS 602
QY 105 QWNAKNSDI--TVQYQGGNNAAL-----VNQTSADSSVMVQVGFNGNAT 147
DQ 603 LINAQTLTISGTIGIIGANNITLQGFNIGSSKTLNGGNVAINELVIGNGS 654

RESULT 11
ICEN PANAN STANDARD; PRT; 1034 AA.
AC Q47879;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ice nucleation protein inau.
GN INAU.
OS Pantoea ananas (Erwinia uredovora).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pantoea.
OX NCBI_TaxID=553;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=KUIN-3;
RA MEDLINE=94264407; PubMed=7764866;
RA Michigami Y., Watabe S., Abe K., Obata H., Arai S.;
RT "Cloning and sequencing of an ice nucleation active gene of Erwinia
RT uredovora."
RL Biosci. Biotechnol. Biochem. 58:762-764(1994).
CC -/- FUNCTION: Ice nucleation proteins enable bacteria to nucleate
CC crystallization in supercooled water.
CC -/- SUBCELLULAR LOCATION: Outer membrane.
CC -/- DOMAIN: CONTAINS IMPERFECT REPEATS OF A CONSENSUS OCTAPEPTIDE
CC A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A REGIONAL 48-RESIDUE
CC PERIODICITY IS SUPERIMPOSED.
CC -/- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE
CC NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.
CC -/- SIMILARITY: Belongs to the bacterial ice nucleation protein
CC family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D14982; BAA03636.1; -.
CC FIR; JC2143; JC2143.
CC HSP; P06620; IINA.
CC InterPro; IPR000258; Ice nucleatn.
CC Pfam; PF00818; Ice nucleation; 51.
CC PRINTS; PR00327; ICNUCLEATN.
CC PROSITE; PS00314; ICE NUCLEATION; 34.
CC Ice nucleation; Repeat; Outer membrane.
CC DOMAIN 162 993 OCTAPEPTIDE PERIODICITY.
CC SEQUENCE 1034 AA; 103378 MW; FA222523D333EADD CRC64;
CC -----
CC Query Match 12.9%; Score 97.5; DB 1; Length 1034;
CC Best Local Similarity 28.7%; Pred. No. 0.85;
CC Matches 43; Conservative 24; Mismatches 46; Indels 37; Gaps 11;

QY 12 IVVSGSALGVYDQLVTRVTHVMAHASGPDSTLSIYQXSANAALALQSDARKSETTI 70
DQ 161 IATYGSITLGSQSLIAGYGSTETA---GDSSTL-IAGYGSTGTA-----GSDSTL 208

```

```

QY 71 TQSGYNGADVGGADNSTIETQNGFRNNATIDQWNAKNSDITVQCYG-----GNNAAL 125
DQ 209 V-AGYGSTQTAGESSQMA-----GYGSTQT-----GMRGSDLTAG-YGSTGTAGDSSSL 256
QY 126 V-----NQTASDSSVMVQVGFNGNATANQ 150
DQ 257 IAGYGSTQTAGESSLT--AGYGSTQTATQK 284

RESULT 12
GP63 LEIMA STANDARD; PRT; 602 AA.
AC P08148; P15906;
DT 01-AUG-1988 (Rel. 08, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Leishmanolysin precursor (EC 3.4.24.36) (Cell surface protease)
DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface
DE endopeptidase).
GN GP63.
OS Leishmania major.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RC SEQUENCE FROM N.A., AND SEQUENCE OF 101-123.
RC MEDLINE=88154764; PubMed=3346625;
RA Button L.L., McMaster W.R.;
RT "Molecular cloning of the major surface antigen of leishmania."
RL J. Exp. Med. 167:724-729(1988).
RN [2]
RC REVISIONS.
RA Button L.L., McMaster W.R.;
RL J. Exp. Med. 171:589-589(1990).
RN [3]
RC GPI-ANCHOR.
RC MEDLINE=91009116; PubMed=2145267;
RA Schneider P., Ferguson M.A.J., McConville M.J., Mehler A.,
RA Romans S.W., Bordier C.;
RT "Structure of the glycosyl-phosphatidylinositol membrane anchor of
RT the Leishmania major promastigote surface protease."
RL J. Biol. Chem. 265:16355-16364(1990).
RN [4]
RC X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).
RC MEDLINE=95406217; PubMed=7675788;
RA Schlagenhauf E., Etges R., Metcalf P.;
RT "Crystallization and preliminary X-ray diffraction studies of
RT leishmanolysin, the major surface metalloproteinase from Leishmania
RT major."
RL Proteins 22:58-66(1995).
RN [5]
RC X-RAY CRYSTALLOGRAPHY (1.86 ANGSTROMS).
RC MEDLINE=98416698; PubMed=9739094;
RA Schlagenhauf E., Etges R., Metcalf P.;
RT "The crystal structure of the Leishmania major surface proteinase
RT leishmanolysin."
RL Structure 6:1035-1046(1998).
CC -/- FUNCTION: Has an integral role during the infection of macrophages
CC in the mammalian host.
CC -/- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and
CC P1', and basic residues at P2 and P3'. A model nonapeptide is
CC cleaved at -Ala-Tyr-[Leu-Lys-Lys-.
CC -/- COFACTOR: Binds 1 zinc ion per subunit.
CC -/- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -/- FULLY SATURATED, UNBRANCHED 1-O-ALKYL CHAIN (MAINLY C24:0) AND A
CC MIXTURE OF FULLY SATURATED UNBRANCHED 2-O-ACYL CHAINS (C12:0,
CC C14:0, C16:0, AND C18:0).
CC -/- SIMILARITY: Belongs to peptidase family M8.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.ieb-sib.ch/announce/> or send an email to license@sib-sib.ch).

```

-----
EMBL; Y00647; CAA68673.1; -
PIR; PLO221; PLO221.
PDB; 1LML; 17-SEP-97.
DR MEROPS; M08.001; -.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR InterPro; IPR001577; Peptidase_M8.
DR Pfam; PF01457; Peptidase_M8; 1.
DR PRINTS; PR00782; LSHMANOLYSIN.
KW Hydrolase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;
  Zymogen; Signal; Cell adhesion; 3D-structure; Lipoprotein.
FT SIGNAL 1 39
FT PROPEP 40 100
FT CHAIN 101 577
FT PROPEP 578 602
FT METAL 264 264
FT ACT_SITE 265 265
FT METAL 268 268
FT METAL 334 334
FT DISULFID 125 142
FT DISULFID 191 230
FT DISULFID 314 386
FT DISULFID 393 455
FT DISULFID 406 425
FT DISULFID 415 489
FT DISULFID 466 510
FT DISULFID 515 565
FT DISULFID 535 558
FT CARBOHYD 300 300
FT CARBOHYD 407 407
FT LIPID 577 577
FT STRAND 101 102
FT STRAND 107 108
FT STRAND 111 114
FT HELIX 116 119
FT TURN 121 122
FT TURN 128 129
FT STRAND 131 133
FT STRAND 139 141
FT HELIX 144 146
FT HELIX 150 158
FT TURN 159 159
FT HELIX 160 169
FT TURN 170 171
FT STRAND 172 174
FT STRAND 177 178
FT STRAND 180 181
FT TURN 189 190
FT HELIX 191 193
FT HELIX 198 202
FT TURN 203 203
FT STRAND 205 206
FT STRAND 210 215
FT TURN 221 222
FT STRAND 226 232
FT TURN 234 235
FT STRAND 238 244
FT HELIX 247 249
FT HELIX 256 269
FT TURN 270 271
FT HELIX 274 279
FT TURN 280 281
FT STRAND 283 286
FT HELIX 289 291
FT STRAND 296 299
FT STRAND 302 312
FT TURN 313 313
FT TURN 315 316
FT STRAND 320 322
N-LINKED (GLCNAC. . .) (POTENTIAL).
N-LINKED (GLCNAC. . .) (POTENTIAL).
GPI-anchor amidated asparagine.

```

```

FT TURN 328 332
FT STRAND 335 335
FT TURN 337 339
FT TURN 341 342
FT STRAND 343 343
FT TURN 344 345
FT STRAND 353 353
FT STRAND 356 364
FT HELIX 365 366
FT TURN 369 370
FT STRAND 372 374
FT TURN 380 383
FT HELIX 386 390
FT STRAND 394 395
FT TURN 396 397
FT STRAND 398 399
FT TURN 402 404
FT STRAND 413 414
FT TURN 417 418
FT STRAND 421 425
FT STRAND 428 429
FT HELIX 435 437
FT TURN 443 444
FT STRAND 445 446
FT TURN 450 454
FT STRAND 458 465
FT TURN 466 467
FT HELIX 470 472
FT TURN 475 477
FT HELIX 478 480
FT TURN 485 486
FT STRAND 487 494
FT STRAND 496 496
FT STRAND 506 516
FT TURN 517 520
FT STRAND 521 525
FT TURN 527 528
FT STRAND 533 534
FT TURN 537 538
FT STRAND 540 542
FT HELIX 543 545
FT TURN 546 546
FT STRAND 550 550
FT TURN 552 553
FT STRAND 555 557
FT HELIX 561 565
FT TURN 566 567
FT HELIX 569 572
FT TURN 573 573
SQ SEQUENCE 602 AA; 63953 MW; 982EF3245D87C43E CRC64;
Query Match 12.7%; Score 96.5; DB 1; Length 602;
Best Local Similarity 59.5%; Pred. NO. 0.56;
Matches 22; Conservative 3; Mismatches 9; Indels 3; Gaps 1;
QY 8 AFPAIVVGSALAGVYDQLVTRVVTHEMAHA--SQP 41
Db 239 AVGVINIPANIASRYDQLVTRVVTHEMAHALGFSGF 275
RESULT 13
ICEN ERWHE STANDARD; PRT; 1258 AA.
AC P16239;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ice nucleation protein.
GN ICEE.
OS Erwinia herbicola.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pantoea.
OX NCBI_TaxID=549;

```

```

RN SEQUENCE FROM N.A.
RP STRAIN=M1;
RX MEDLINE=90152370; PubMed=2515997;
RA Warren G.J., Corotto L.V.;
RT "The consensus sequence of ice nucleation proteins from Erwinia
RT herbicola, Pseudomonas fluorescens and Pseudomonas syringae.";
RL Gene 85:239-242(1989).
CC -!- FUNCTION: Ice nucleation proteins enable bacteria to nucleate
CC -!- crystallization in supercooled water.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- DOMAIN: CONTAINS 126 IMPERFECT REPEATS OF A CONSENSUS OCTAPEPTIDE
CC A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A REGIONAL 48-RESIDUE
CC PERIODICITY IS SUPERIMPOSED.
CC -!- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE
CC NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.
CC -!- SIMILARITY: Belongs to the bacterial ice nucleation protein
CC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; M26382; AAA24823.1; -
CC DR PIR; JQ0188; JQ0188.
CC DR HSSP; P06620; IINA.
CC DR InterPro; IPR000258; Ice nucleatn.
CC DR Pfam; PF00818; Ice nucleation; 65.
CC DR PRINTS; PR00327; ICENUCLEATN.
CC DR PROSITE; PS00314; ICE_NUCLEATION; 45.
CC KW Ice nucleation; Repeat; Outer membrane.
CC FT DOMAIN 162 1217 OCTAPEPTIDE PERIODICITY.
CC SQ SEQUENCE 1258 AA; 125084 MW; 590EBA13007FBD4 CRC64;

Query Match 12.6%; Score 95.5; DB 1; Length 1258;
Best Local Similarity 28.7%; Pred. No. 1.6;
Matches 43; Conservative 24; Mismatches 46; Indels 37; Gaps 11;

QY 12 IVVSGSALAGVY-DQLVTRVVTWTHMAHSGPDSTLSIYQYGSANAALQSDARKSETTI 70
Db 161 IATYGSTLSTGTHQSOLIAGYGSTETA---GDSSTL-IAGYGSTGTAGA-----DSTL 208

QY 71 TQSGYNGADVQGGADNSTIETQNGFRNATIDOWNAKNSDITVGYG-----GNRAAL 125
Db 209 V-AGYGSTQTAGESSQMA-----GYGSTQ-----GMKGSDLTAG-YGSTGTAGDSSSL 256

QY 126 V-----NOTASDSVMVRQVGFNNATANQ 150
Db 257 IAGYGSTQTAGEDSSLT---AGYGSTQTAK 284

RESULT 14
ID ICEA PANAN STANDARD; PRT; 1322 AA.
AC P20469;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ice nucleation protein inaA.
GN INAA.
OS Pantoea ananas (Erwinia uredovora).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pantoea.
OX NCBI_TaxID=553;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90092494; PubMed=2599095;
RA Abe K., Watabe S., Emori Y., Watanabe M., Arai S.;
RT "An ice nucleation active gene of Erwinia ananas. Sequence similarity

```

```

RT to those of Pseudomonas species and regions required for ice
RT nucleation activity."
RL FEBS Lett. 258:297-300(1989).
CC -!- FUNCTION: Ice nucleation proteins enable bacteria to nucleate
CC crystallization in supercooled water.
CC -!- SUBCELLULAR LOCATION: Outer membrane (By similarity).
CC -!- DOMAIN: CONTAINS MANY IMPERFECT REPEATS OF THE CONSENSUS
CC OCTAPEPTIDE A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A
CC REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.
CC -!- SIMILARITY: Belongs to the bacterial ice nucleation protein
CC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; X17316; CAA35194.1; -
CC DR PIR; S07053; S07053.
CC DR HSSP; P06620; IINA.
CC DR InterPro; IPR000258; Ice nucleatn.
CC DR Pfam; PF00818; Ice nucleation; 69.
CC DR PRINTS; PR00327; ICENUCLEATN.
CC DR PROSITE; PS00314; ICE_NUCLEATION; 49.
CC KW Ice nucleation; Repeat; Outer membrane.
CC FT DOMAIN 162 1281 OCTAPEPTIDE PERIODICITY.
CC SQ SEQUENCE 1322 AA; 131094 MW; 89B0BE24AA837039 CRC64;

Query Match 12.5%; Score 94.5; DB 1; Length 1322;
Best Local Similarity 29.3%; Pred. No. 2.1;
Matches 41; Conservative 20; Mismatches 46; Indels 33; Gaps 10;

QY 12 IVVSGSALAGVY-DQLVTRVVTWTHMAHSGPDSTLSIYQYGSANAALQSDARKSETTI 70
Db 161 IATYGSTLSTGTHQSOLIAGYGSTETA---GDSSTL-IAGYGSTGTAGA-----GSDSTL 208

QY 71 TQSGYNGADVQGGADNSTIETQNGFRNATIDOWNAKNSDITVGYGNNALVNOTA 130
Db 209 V-AGYGSTQTAGESSQMA-----GYGSTQ-----GMKGSDLTAG-YGSTGTAGA-----G 251

QY 131 SDSVMVRQVGFNNATANQ 150
Db 252 ADSLSI---AGYGSTQTAGE 268

RESULT 15
ID ICEN XANCT STANDARD; PRT; 1567 AA.
AC P18127;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ice nucleation protein.
GN INAX.
OS Xanthomonas campestris (pv. translucens).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=343;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=X56S;
RX MEDLINE=91080859; PubMed=2259339;
RA Zhao J., Orser C.S.;
RT "Conserved repetition in the ice nucleation gene inax from
RT Xanthomonas campestris pv. translucens.";
RL Mol. Gen. Genet. 223:163-166(1990).
CC -!- FUNCTION: Ice nucleation proteins enable bacteria to nucleate
CC crystallization in supercooled water.
CC -!- SUBCELLULAR LOCATION: Outer membrane (By similarity).
CC -!- DOMAIN: CONTAINS 153 IMPERFECT REPEATS OF THE CONSENSUS

```



```
CC OCTAPEPTIDE A-G-Y-G-S-T-L-T; FURTHER ON A 16-RESIDUE AND A
CC REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.
CC -!- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE
CC NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.
CC -!- SIMILARITY: Belongs to the bacterial ice nucleation protein
CC family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X52970; CAA37140.1; -.
DR HSP; P06620; IINA.
DR InterPro; IPR000258; Ice_nucleatn.
DR Pfam; PF00818; Ice_nucleatn; 81.
DR PRINTS; PR00327; ICENUCLEATN.
DR PROSITE; PS00314; ICE_NUCLEATION; 57.
KW Ice_nucleation; Repeat; Outer membrane.
SQ SEQUENCE 1567 AA; 152548 MW; C8B451D959ECAD63 CRC64;
Query Match 12.4%; Score 93.5; DB 1; Length 1567;
Best Local Similarity 29.9%; Pred. No. 3;
Matches 43; Conservative 26; Mismatches 30; Indels 45; Gaps 12;
QY 39 SGPDSLTSLIYQYGSANAALQSDARKSETTITSGYGN-----GADV-----GQG 84
DB 1060 AGADSTL-IAGYGSTQTA-----GSDSLT-AGYGTQTARQGSITAGYGTGTAG 1109
QY 85 ADNSTIE---LTQN-GPRNNATI---DOWNAKNSDITVGOYG-----GNNALV----- 126
DB 1110 ADSSLIAGYGTQTAGYDSNLITAGYGTQTARESSLTAG-YGSTSTAGHDSLIAGYGS 1168
QY 127 NQTASDSSVMVRQVGFNNATANQ 150
DB 1169 TQTAGYNSILT--TGYGSTQTAQE 1190
```

Search completed: August 2, 2004, 14:49:29
Job time : 6.3 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:39:17 ; Search time 29.7 Seconds

(without alignments)
1604.150 Million cell updates/sec

Title: US-09-543-407-16

Perfect score: 757
Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVFGNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL 25:*

- 1: sp_arChaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriap:*
- 17: sp_arChaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	649	85.7	152	2 Q33802	Q33802 salmonella
2	563.5	74.4	150	2 Q7X243	Q7X243 citrobacter
3	524	69.2	149	2 Q7X240	Q7X240 citrobacter
4	471.5	62.3	152	16 Q8CW63	Q8CW63 escherichia
5	454.5	60.0	150	2 Q7X237	Q7X237 enterobacte
6	340	44.9	76	2 Q54069	Q54069 salmonella
7	133	17.6	502	16 Q8EIH4	Q8EIH4 shewanella
8	120	15.9	160	16 Q8CW64	Q8CW64 escherichia
9	120	15.9	160	16 Q83R07	Q83R07 shigella fl
10	118.5	15.7	151	16 Q7UCZ1	Q7UCZ1 shigella fl
11	118.5	15.7	153	16 Q89JI6	Q89JI6 bradyrhizob
12	116	15.3	171	16 Q89JI3	Q89JI3 bradyrhizob
13	115	15.2	139	16 Q8EIH3	Q8EIH3 shewanella
14	110	14.5	130	16 Q89JI4	Q89JI4 bradyrhizob
15	108.5	14.3	151	2 Q7X244	Q7X244 citrobacter
16	107.5	14.2	154	16 Q89JI5	Q89JI5 bradyrhizob

17	105.5	13.9	91	2 Q9S3J8	Q9S3J8 escherichia
18	102.5	13.5	152	2 Q7X241	Q7X241 citrobacter
19	101.5	13.4	2087	5 Q8MVW7	Q8MVW7 naegleria g
20	101	13.3	599	5 Q25273	Q25273 leishmania
21	99.5	13.1	582	16 P71868	P71868 mycobacteri
22	99.5	13.1	582	16 Q7TW98	Q7TW98 mycobacteri
23	99.5	13.1	641	5 Q8MN20	Q8MN20 leishmania
24	99.5	13.1	657	5 Q8MN21	Q8MN21 leishmania
25	99	13.1	1613	2 Q840U5	Q840U5 rickettsia
26	99	13.1	1618	2 Q7XKB1	Q7XKB1 rickettsia
27	98.5	13.0	151	2 Q7X238	Q7X238 enterobacte
28	98	12.9	145	16 Q8U6N9	Q8U6N9 agrobacte
29	98	12.9	157	16 Q8HGO	Q8HGO pseudomonas
30	97	12.8	1612	2 Q840U6	Q840U6 rickettsia
31	97	12.8	1617	2 Q7X5N9	Q7X5N9 rickettsia
32	96.5	12.7	348	13 Q93397	Q93397 cyprinus ca
33	96.5	12.7	644	5 Q43994	Q43994 leishmania
34	96.5	12.7	1306	2 Q93N36	Q93N36 pantoea ana
35	96.5	12.7	1341	16 Q8ED31	Q8ED31 shewanella
36	96.5	12.7	2016	5 Q9BIT0	Q9BIT0 plectreury
37	96	12.7	598	5 Q25275	Q25275 leishmania
38	96	12.7	641	5 Q8MN22	Q8MN22 leishmania
39	96	12.7	641	5 Q8MN48	Q8MN48 leishmania
40	96	12.7	641	5 Q8MN23	Q8MN23 leishmania
41	96	12.7	641	5 Q8MN24	Q8MN24 leishmania
42	96	12.7	641	5 Q8MN25	Q8MN25 leishmania
43	95.5	12.6	179	2 Q33801	Q33801 salmonella
44	95.5	12.6	617	5 Q94593	Q94593 leishmania
45	95.5	12.6	1613	2 Q9KKB2	Q9KKB2 israeli tic

ALIGNMENTS

RESULT 1

Q33802	PRELIMINARY;	PRT;	152 AA.
ID	Q33802		
AC	Q33802		
DT	01-JAN-1998 (TrEMBLrel. 05, Created)		
DT	01-JAN-1998 (TrEMBLrel. 05, Last sequence update)		
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)		
DE	Agfa protein (fragment).		
GN	AGFA.		
OS	Salmonella typhimurium.		
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;		
OC	Enterobacteriaceae; Salmonella.		
OX	NCBI_TaxID=602;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=98053981; PubMed=9393832;		
RA	Sukupolvi S.S., Lorentz R.G., Gordon J.I., Bian Z., Pfeifer J.D.,		
RA	Normark S.J., Rhen M.;		
RT	"Expression of thin, aggregative fimbriae promotes interaction of		
RT	Salmonella typhimurium SR-11 with mouse small intestinal epithelial		
RT	cells."		
RL	Infect. Immun. 65:5320-5325(1997).		
DR	EMBL; AJ000514; CAA04151.1; -.		
FT	NON TER		
SQ	SEQUENCE	152 AA; 15401 MW; 9DA7DADC2364B006 CRC64;	

Query Match	85.7%;	Score 649; DB 2; Length 152;
Best Local Similarity	88.7%;	Pred. No. 1.1e-44;
Matches 134; Conservative	4; Mismatches 13; Indels	0; Gaps 0;
QY	1	MKLLKVAFAAIVVSGSALAGVVDQLVTRVWTHMAHASGPDSTLSIYVGSANALALQ 60
Db	1	MKLLKVAFAAIVVSGSALAGVVDQLVTRVWTHMAHASGPDSTLSIYVGSANALALQ 60
QY	61	SDARKSETTITGSGYNGADVQGGADNSTIELTQNGFRNNATIDQNAKNSITVQYGG 120
Db	61	SDARKSETTITGSGYNGADVQGGADNSTIELTQNGFRNNATIDQNAKNSITVQYGG 120
QY	121	NNAALVNTQATSDSSVMVRQVFGNNATANQY 151

```

Db      121 NNAALVNQASDSSVMVRQVGFNNATANQY 151
|||||
RESULT 2
Q7X243
ID Q7X243 PRELIMINARY; PRT; 150 AA.
AC Q7X243
DT 01-OCT-2003 (T-EMBLrel. 25, Created)
DT 01-OCT-2003 (T-EMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Curlin-csgA protein.
GN CSGA.
OS Citrobacter sp. Fec2.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Citrobacter.
OX NCBI_TaxID=213763;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fec2;
RA Zogaj X., Bokranz W., Nimitz M., Romling U.;
RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RT Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158(2003).
DR EMBL; AJ515700; CAD56672.1; -.
SQ SEQUENCE 150 AA; 15016 MW; 1D7141B8D6973DC6 CRC64;

Query Match 74.4%; Score 563.5; DB 2; Length 150;
Best Local Similarity 74.2%; Pred. No. 8e-38;
Matches 118; Conservative 11; Mismatches 13; Indels 17; Gaps 2;

QY 1 MKLLKVAAPAAIVVSGSALAGVYDQVTRVWTHM-----AHASGPDSTLSIYQVGS 52
|||||
DB 1 MKLLQVAAPAAIVVSGSALAGVVPQ-----WGGGGGGGGSSSGPESTLSIYQSGV 51
|||||
QY 53 ANAALALQSDARKSETTITQSGYNGADVGQGDNSTIELTQNGFNNATIDQWNAKNSD 112
|||||
DB 52 NNAALALQSDARKSDTTIHQNGFGNGADVGQGSDDSIDLTQRFNGSATLDQWNGKDSMTVTKQFG 119
|||||

RESULT 3
Q7X240
ID Q7X240 PRELIMINARY; PRT; 149 AA.
AC Q7X240
DT 01-OCT-2003 (T-EMBLrel. 25, Created)
DT 01-OCT-2003 (T-EMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Curlin-csgA protein.
GN CSGA.
OS Citrobacter freundii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Citrobacter.
OX NCBI_TaxID=546;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fec4;
RA Zogaj X., Bokranz W., Nimitz M., Romling U.;
RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RT Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158(2003).
DR EMBL; AJ515701; CAD56675.1; -.
SQ SEQUENCE 149 AA; 15260 MW; 946DD52017F648FD CRC64;

Query Match 69.2%; Score 524; DB 2; Length 149;
Best Local Similarity 72.8%; Pred. No. 1.2e-34;
Matches 110; Conservative 14; Mismatches 25; Indels 2; Gaps 1;

QY 1 MKLLKVAAPAAIVVSGSALAGVYDQVTRVWTHMNAHASGPDSTLSIYQVGSANAALAQ 60
|||||

```

```

Db      121 NNAALVNQASDSSVMVRQVGFNNATANQY 151
|||||
RESULT 2
Q7X243
ID Q7X243 PRELIMINARY; PRT; 150 AA.
AC Q7X243
DT 01-OCT-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Major curlin subunit precursor.
GN CSGA OR C1306.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
RC MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D.A., Burland V., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
DR EMBL; AE016759; AAN79779.1; -.
KW Complete proteome.
SQ SEQUENCE 152 AA; 15064 MW; 3BA57F34C1240E83 CRC64;

Query Match 62.3%; Score 471.5; DB 16; Length 152;
Best Local Similarity 65.8%; Pred. No. 2e-30;
Matches 100; Conservative 18; Mismatches 33; Indels 1; Gaps 1;

QY 1 MKLLKVAAPAAIVVSGSALAGVYDQVTRVWTH-ENAHASGPDSTLSIYQVGSANAALAQ 59
|||||
DB 1 MKLLKVAAPAAIVVSGSALAGVVPQVGCGGNGHGGGNGNSSELSNIYQGGNSALAQ 60
|||||
QY 60 QSDARKSETTITQSGYNGADVGQGDNSTIELTQNGFNNATIDQWNAKNSDITVQYQG 119
|||||
DB 61 QADARNSDLTITQHGCGNGADVGQGSDDSIDLTQRFNGSATLDQWNGKDSMTVTKQFG 120
|||||
QY 120 GNNALVNQASDSSVMVRQVGFNNATANQY 151
|||||
DB 121 GNGAAVDQTSNNSVNVTVQVGFNNATANQY 152
|||||

RESULT 5
Q7X237
ID Q7X237 PRELIMINARY; PRT; 150 AA.
AC Q7X237
DT 01-OCT-2003 (T-EMBLrel. 25, Created)
DT 01-OCT-2003 (T-EMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Curlin-csgA protein.
GN CSGA.
OS Enterobacter sakazakii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Enterobacter.
OX NCBI_TaxID=28141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fec39;
RA Zogaj X., Bokranz W., Nimitz M., Romling U.;

```

```

RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RL Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158(2003).
DR EMBL: AJ515702; CAD56678.1; -.
SQ SEQUENCE 150 AA; 15112 MW; 508BB2D872DF15F3 CRC64;

Query Match 60.0%; Score 454.5; DB 2; Length 150;
Best Local Similarity 62.1%; Pred. No. 4.6e-29;
Matches 95; Conservative 24; Mismatches 29; Indels 5; Gaps 2;

QY 1 MKLKVAAFAAIVVSGSALAGVYDQVLTFRVVVTHEMAHA--SGPDSLTIYQYGSANAALA 58
DB 1 MKFKVAAALAIIVVSGSAMAGVNIQ---GGWGHGGHGGYGGPSTLNIIYQNGGNSALA 57

QY 59 LOSDARKSETTITQSGYNGADVGCGADNSTIETQNGFRNATIDQNAKNSDITVQY 118
DB 58 LQTDARNVNLISOTGGGNGADVGCGDSSINITONGFGNSATLDQWNSKDSVMNVSY 117

QY 119 GGNNAALVNQNTASDSSVMVROVGGNNATANQY 151
DB 118 GGLNGALVDQASNSVNTVQIGFGNHATAHQY 150

RESULT 6
Q54069 PRELIMINARY; PRT; 76 AA.
AC Q54069 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE SEF17 fimbria (Fragment).
GN AGFA.
OS Salmonella enteritidis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=592;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SE30;
RA Cox J.M., Eglezos S., Woolcock J.B.;
RT "Virulence of Salmonella enteritidis in chickens correlates with
RL colony morphology and expression of SEF17 fimbriae.";
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: U53207; AAA98671.1; -.
FT NON_TER 1
FT NON_TER 76
FT NON_TER 76
SQ SEQUENCE 76 AA; 7704 MW; 2FD5411241A7BCB1 CRC64;

Query Match 44.9%; Score 340; DB 2; Length 76;
Best Local Similarity 97.1%; Pred. No. 3.1e-20;
Matches 67; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 37 HASGPDSTLSIYQGSANAALQSDARKSETTITQSGYNGADVGCGADNSTIETQNG 96
DB 8 NSSGPDSTLSIYQGSANAALQSDARKSETTITQSGYNGADVGCGADNSTIETQNG 67

QY 97 FRNNATIDQ 105
DB 68 FRNNATIDQ 76

RESULT 7
Q8EIH4 PRELIMINARY; PRT; 502 AA.
AC Q8EIH4 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DE 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Conserved hypothetical protein.
GN SO0865.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.

```

```

OX NCBI_TaxID=70863;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NR-1;
RX MEDLINE=22297686; PubMed=12368813;
RA Keidelberg J.F., Paulsen I.T., Neilson K.E., Gaidos E.J., Neilson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
RA Meyer T., Isapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
RA Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,
RA Mueller J., Khouri H., Gill J., Uterback T.R., McDonald L.A., C.M.;
RA Feldblum T.V., Smith H.O., Venter J.C., Neilson K.H., Fraser C.M.;
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis.";
RL Nat. Biotechnol. 20:1118-1123(2002).
DR EMBL: AE015332; AAN53941.1; -.
DR TIGR: SO0865; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 502 AA; 52441 MW; D08CA23D6C46B62D CRC64;

Query Match 17.6%; Score 133; DB 16; Length 502;
Best Local Similarity 29.6%; Pred. No. 0.013;
Matches 37; Conservative 21; Mismatches 59; Indels 8; Gaps 3;

QY 32 THEMAHAGPDSSTLSIYQYGSANA---ALALQSDARKSETTITQSGYNGADVGCGADNS 88
DB 258 TAYLSMTTGGDNDVDTQGDNSTVGVDSLIADIQGDNDNITIKQKDSNGAEFQVWGDSN 317

QY 89 TIELTQNGFRNATIDQNAKNSDITVQYGGNNAALVNQNTASDSSVMVROVQFGN---- 144
DB 318 DVDLQKQGDANFATFGAYGTDN--DFDLSKGDNNELVAFATGSDNSIEISQEGDANFAYV 376

QY 145 NATAN 149
DB 377 DATGN 381

RESULT 8
Q8CW64 PRELIMINARY; PRT; 160 AA.
AC Q8CW64 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DE 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Minor curlin subunit precursor.
GN CSGB OR C1305.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G., III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Rackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
DR EMBL: AE016759; AAN79778.1; -.
DX Complete proteome.
SQ SEQUENCE 160 AA; 16963 MW; 49F68448D979B986 CRC64;

Query Match 15.9%; Score 120; DB 16; Length 160;
Best Local Similarity 28.7%; Pred. No. 0.036;
Matches 41; Conservative 17; Mismatches 67; Indels 18; Gaps 3;

QY 22 VYDQ-----LTVRVVTHEMAHAGSPDSTLSIYQYGSANAALQSDARKSET 69
DB 1 MYDQVQGDNMKNKLLFMMLTILGAPGIAAAGYDLANSEYNF-----AVNELSKSFPQAA 56

```

QY 69 TITQSGYNGADYVGQADNSTDITLQNGFRNNATIDQNAKNSDITVQYGGNNAALVNO 128
 DB 57 IICQAGTNNNSAQRQGGSKLLTVVAGGSSNRKAKIDQTDGYNL-AYIDQAGSANDASISQ 115
 QY 129 TASDSSVMVRQVGFNNATANQY 151
 DB 116 GAYGNTAMIIQKSGNKNITQY 138

RESULT 9

Q83RU7 PRELIMINARY; PRT; 160 AA.
 AC Q83RU7;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Minor curlin subunit precursor, similar to CsgA.
 GN CSGB OR SF1035.
 OS Shigella flexneri.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Shigella.
 OX NCBI_TaxID=623;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=301 / Serotype 2a;
 RX MEDLINE=22272406; PubMed=12384590;
 RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
 RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
 RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
 RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
 RA Yu J.;
 RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
 RT through comparison with genomes of Escherichia coli K12 and O157."
 RL Nucleic Acids Res. 30:4432-4441(2002).
 RL EMBL; A5015131; AAN42658.1; -
 KW Complete proteome.
 SQ SEQUENCE 160 AA; 16919 MW; 50269F5268D2A32F CRC64;

Query Match 15.9%; Score 120; DB 16; Length 160;
 Best Local Similarity 28.7%; Pred. No. 0.036;
 Matches 41; Conservative 17; Mismatches 67; Indels 18; Gaps 3;
 QY 22 VVDQ-----LVTRVVTHEMAHASGPDSTLSIYQYGSANAALAQSDARKSET 68
 DB 1 MYDQVGDNMKNKLLFMMLTILGAGIAAAGYDLANSEYNF----AVNELSKSFSNQAA 56
 QY 69 TITQSGYNGADYVGQADNSTDITLQNGFRNNATIDQNAKNSDITVQYGGNNAALVNO 128
 DB 57 IICQAGTNNNSAQRQGGSKLLAVVAGGSSNRKAKIDQTDGYNL-AYIDQAGSANDASISQ 115
 QY 129 TASDSSVMVRQVGFNNATANQY 151
 DB 116 GAYGNTAMIIQKSGNKNITQY 138

RESULT 10

Q7UCZ1 PRELIMINARY; PRT; 151 AA.
 AC Q7UCZ1;
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Minor curlin subunit.
 GN CSGB OR S1108.
 OS Shigella flexneri.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Shigella.
 OX NCBI_TaxID=623;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=2457T / ATCC 700930 / Serotype 2a;
 RX MEDLINE=22590274; PubMed=12704152;

RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
 RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
 RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
 RA Schwartz D.C., Blattner F.R.;
 RT "Complete genome sequence and comparative genomics of Shigella
 RT flexneri serotype 2a strain 2457T."
 RL Infect. Immun. 71:2775-2786(2003).
 RL EMBL; AE016981; AAP16542.1; -
 SQ SEQUENCE 151 AA; 15868 MW; 5D5D266B964014A0 CRC64;

Query Match 15.7%; Score 118.5; DB 16; Length 151;
 Best Local Similarity 31.6%; Pred. No. 0.045;
 Matches 37; Conservative 13; Mismatches 62; Indels 5; Gaps 2;

QY 35 MAHASGPDSTLSIYQYGSANAALQSDARKSETTITQSGYNGADVQAGADNSTIELTQ 94
 DB 18 IAAAGYDLANSEYNF----AVNELSKSFSNQAAIICQAGTNNNSAQRQGGSKLLAVVQ 73
 QY 95 NGFRNNATIDQNAKNSDITVQYGGNNAALVNQTSASDSVMVRQVGFNNATANQY 151
 DB 74 EGSSNRKAKIDQTDGYNL-AYIDQAGSANDASISQAGVGNNTAMIIQKSGNKNITQY 129

RESULT 11

Q89JI6 PRELIMINARY; PRT; 153 AA.
 AC Q89JI6;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE CSGB protein.
 GN CSGB OR BLIS297.
 OS Bradyrhizobium japonicum.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Bradyrhizobiaceae; Bradyrhizobium.
 OX NCBI_TaxID=375;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=USDA 110;
 RX MEDLINE=22484998; PubMed=12597275;
 RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiuni T.,
 RA Sasamoto S., Watanabe A., Idesawa K., Iriguchi M., Kawashima K.,
 RA Kohara M., Matsumoto M., Shimo S., Tsuruoka H., Wada T., Yamada M.,
 RA Tabata S.;
 RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
 RT Bradyrhizobium japonicum USDA110."
 RL DNA Res. 9:189-197(2002).
 RL EMBL; AP005954; BAC50562.1; -
 KW Complete proteome.
 SQ SEQUENCE 153 AA; 15991 MW; 4CE71DEAC375145B CRC64;

Query Match 15.7%; Score 118.5; DB 16; Length 153;
 Best Local Similarity 27.0%; Pred. No. 0.046;
 Matches 41; Conservative 32; Mismatches 56; Indels 23; Gaps 6;

QY 2 KLIKVA--FAAIVSGSALAGYVDQLVTRVTHEMAHASGPDSTL-SIYQYGSANAAL 59
 DB 10 RVLVAALLAAGCAATQASAGSIQSVT-----NPNVSIETIVQFQNDVPVTI 58
 QY 60 QSDARKSETTITQSGYNGADVQAGADNSTDITLQNGFRNNATIDQNAKNSDITVQYQ 119
 DB 59 EENSRYNIARVQIG-----GSGTVDATI--IQNGTRNYANVIMGG-TTNAAGQSG 108
 QY 120 GNNALVNQTSASDSVMVRQVGFNNATANQY 151
 DB 109 LNSTADITQIGNSTNALLIQIDMNSGAVRQF 140

RESULT 12

Q89JI3 PRELIMINARY; PRT; 171 AA.
 AC Q89JI3;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)

Query Match	15.2%;	Score 115;	DB 16;	Length 139;	
Best Local Similarity	27.0%;	Pred. No. 0.077;			
Matches 33;	Conservative	23;	Mismatches 52;	Indels 14;	Gaps 3;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:35:42 ; Search time 44.9 Seconds
(without alignments)
950.215 Million cell updates/sec

Title: US-09-543-407-18
Perfect score: 780
Sequence: 1 MKLKVAFAPAAIVVSGSALA.....DSSVMVRFQVGFNNATANYQ 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1596107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1596107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04.*
1: geneseqp1980s.*
2: geneseqp1990s.*
3: geneseqp2000s.*
4: geneseqp2001s.*
5: geneseqp2002s.*
6: geneseqp2003as.*
7: geneseqp2003bs.*
8: geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	780	100.0	151	3	AAB36349	AAB36349 Agfa::PT3
2	691	88.6	151	2	AAR74625	Aar74625 Agfa sequ
3	691	88.6	151	3	AAB36341	AAB36341 Salmonell
4	686	87.9	151	2	AAW23570	AAW23570 Salmonell
5	675	86.5	151	3	AAB36353	AAB36353 Agfa::PT3
6	622	79.7	151	3	AAB36350	AAB36350 Agfa::PT3
7	618	79.2	151	3	AAB36348	AAB36348 Agfa::PT3
8	613	78.6	151	3	AAB36346	AAB36346 Agfa::PT3
9	611	78.3	151	3	AAB36347	AAB36347 Agfa::PT3
10	608	77.9	151	3	AAB36352	AAB36352 Agfa::PT3
11	601	77.1	151	3	AAB36354	AAB36354 Agfa::PT3
12	600	76.9	151	3	AAB36351	AAB36351 Agfa::PT3
13	599	76.8	151	3	AAB36355	AAB36355 Agfa::PT3
14	523	67.1	151	3	AAB36343	AAB36343 Escherich
15	518	66.4	151	7	ABR82651	ABR82651 E. coli C
16	506	64.9	120	2	AAR62761	Aar62761 Agfa sequ
17	506	64.9	120	2	AAW23569	AAW23569 Salmonell
18	445	57.1	142	2	AAR52664	Aar52664 Fibronect
19	372	47.8	122	2	AAR52663	Aar52663 FNB curli
20	277	30.4	45	3	AAB36316	AAB36316 Salmonell
21	132	16.9	22	3	AAB36318	AAB36318 Salmonell
22	123	15.8	23	3	AAB36321	AAB36321 Salmonell
23	123	15.8	23	3	AAB36326	AAB36326 Salmonell
24	123	15.8	23	3	AAB36338	AAB36338 Salmonell
25	115	14.7	22	3	AAB36325	AAB36325 Salmonell

26	115	14.7	22	3	AAB36339	Aab36339 Salmonell
27	115	14.7	22	3	AAB36320	Aab36320 Salmonell
28	111	14.2	22	3	AAB36322	Aab36322 Salmonell
29	111	14.2	22	3	AAB36327	Aab36327 Salmonell
30	111	14.2	22	3	AAB36337	Aab36337 Salmonell
31	109	14.0	24	7	ABR82644	ABR82644 E. coli c
32	107.5	13.8	151	3	AAB36344	Aab36344 Escherich
33	106.5	13.7	151	3	AAB36342	Aab36342 Salmonell
34	102	13.1	26	7	ABR82649	ABR82649 E. coli V
35	101	12.9	262	4	ABR866342	ABR866342 Drosophil
36	98	12.6	26	7	ABR82645	ABR82645 E. coli c
37	96	12.3	19	3	AAB36323	Aab36323 Salmonell
38	96	12.3	19	3	AAB36336	Aab36336 Salmonell
39	96	12.3	19	3	AAB36328	Aab36328 Salmonell
40	95	12.2	24	7	ABR82647	ABR82647 E. coli c
41	93.5	12.0	287	4	ABR866343	ABR866343 Drosophil
42	93	11.9	974	6	ABU17075	Abu17075 Protein e
43	92	11.8	23	3	AAB36331	Aab36331 Escherich
44	92	11.8	975	6	ADA33477	Ada33477 Acinetoba
45	91.5	11.7	580	5	AAG66008	Aag66008 F. necrop

ALIGNMENTS

RESULT 1
AAB36349 standard; protein; 151 AA.
XX AAB36349;
XX AC AAB36349;
XX DT 26-FEB-2001 (first entry)
XX DE Agfa::PT3#4 amino acid sequence SEQ ID NO:18.
XX KW Salmonella; agfa; chromosomal gene replacement; fimbriin; epitope;
XX KM vaccine; immune response; immunogen.
XX OS Salmonella enteritidis.
XX OS Escherichia coli.
XX OS Synthetic.
XX PN WO200060102-A2.
XX PD 12-OCT-2000.
XX PF 05-APR-2000; 2000WO-CA000356.
XX PR 05-APR-1999; 99US-0127888P.
XX PA (UYVI-) UNIV VICTORIA.
XX PI White AP, Doran JL, Collison SK, Kay WW;
XX PT WPI; 2000-672631/65.
XX DR N-PSDB; AAC64625.
XX PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence which encodes foreign epitope or antigen, expresses recombinant Agfa protein useful for eliciting immune response in animal.
XX PS Disclosure; Page 136; 139pp; English.
XX CC The present invention describes a recombinant agfa gene (I) where a segment of the gene has been replaced by a segment of a foreign DNA sequence which encodes a foreign epitope or antigen. Also described are: (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended assembly system of strains of Salmonella, Escherichia coli and Enterobacteriaceae for the production of fimbriae comprising recombinant Agfa, CsgA and Agfa-homologue fimbriin subunits, respectively; (2) CC directing recombination of a recombinant gene into the chromosome of the homologous species; (3) directing recombination of a recombinant gene back into the chromosome of the homologous species, replacing the native

CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;

Query Match 100.0%; Score 780; DB 3; Length 151;
 Best Local Similarity 100.0%; Pred. No. 1.7e-67; Indels 0; Gaps 0;
 Matches 151; Conservative 0; Mismatches 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGDPDYDQLVTRVVTHEMAHALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGDPDYDQLVTRVVTHEMAHALQ 60
 QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITVGQYGG 120
 DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITVGQYGG 120
 QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
 DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 2
 AAR74625
 ID AAR74625 standard; protein; 151 AA.
 AC AAR74625;
 DT 25-MAR-2003 (revised)
 DT 26-JUN-1995 (first entry)
 XX Agfa sequence.
 DE
 XX Salmonella; Agfa; vaccine.
 KW Salmonella.
 OS
 XX WO9425598-A2.
 PN
 XX 10-NOV-1994.
 PD
 XX 26-APR-1994; 94WO-IB000207.
 PF
 XX 26-APR-1993; 93US-00054452.
 PR
 XX {UYVI-} UNIV VICTORIA INNOVATION & DEV CORP.
 PA (KING/) KING J.
 PI Kay WW, Collinson SK, Clouthier SC, Doran JL;
 XX WPI; 1994-358275/44.
 DR N-PSDB; AAQ87467.
 DR
 XX Eliciting an immune response to Salmonella - using attenuated Salmonella
 PT strains, vector constructs, or compns. contg. fimbrial type proteins.
 PT
 PS Disclosure; Fig 7B; 95pp; English.
 XX
 XX The Salmonella Agfa protein and DNA are used in vaccine and genetic

CC immunization compositions, respectively, to elicit an immune response to
 CC Salmonella in animals (e.g. food producing animals) and humans. (Updated
 CC on 25-MAR-2003 to correct PN field.)
 XX
 SQ Sequence 151 AA;

Query Match 88.6%; Score 691; DB 2; Length 151;
 Best Local Similarity 90.7%; Pred. No. 6.8e-59;
 Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGDPDYDQLVTRVVTHEMAHALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGDPDSTLSIYQYGSANALALQ 60
 QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITVGQYGG 120
 DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITVGQYGG 120
 QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
 DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 3
 AAB36341
 ID AAB36341 standard; protein; 151 AA.
 AC AAB36341;
 XX 26-FEB-2001 (first entry)
 DT
 XX Salmonella enteritidis Agfa amino acid sequence SEQ ID NO:5.
 DE
 XX Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 XX WO200060102-A2.
 PN
 XX 12-OCT-2000.
 PD
 XX 05-APR-2000; 2000WO-CA000356.
 PF
 XX 05-APR-1999; 99US-0127888P.
 PR
 XX {UYVI-} UNIV VICTORIA.
 PA
 PI White AP, Doran JL, Collinson SK, Kay WW;
 XX WPI; 2000-672631/65.
 DR N-PSDB; AAC64617.
 DR
 XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 135; 139pp; English.

CC The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or

CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

XX Sequence 151 AA;

Query Match 88.6%; Score 691; DB 3; Length 151;
 Best Local Similarity 90.7%; Pred. No. 6.8e-59;
 Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

Qy 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGDPDYQLVTRVVTHEMAHALQ 60
 Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGDPSTLSIYQYGSANAALALQ 60
 Qy 61 SDARKSETTITQSGYNGADYVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGOYGG 120
 Db 61 SDARKSETTITQSGYNGADYVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGOYGG 120
 Qy 121 NNAALVNOTASDSSVMVROVGFGNATANQY 151
 Db 121 NNAALVNOTASDSSVMVROVGFGNATANQY 151

RESULT 4

AAW23570
 ID AAW23570 standard; protein; 151 AA.

XX AC AAW23570;

XX 25-MAR-2003 (revised)
 DT 29-SEP-1997 (first entry)

XX Salmonella enteritidis 27655-3b agfa.

XX Enteropathogenic bacteria; enterobacteria; S. enteritidis; antibody.

XX Salmonella enteritidis.

XX Key Location/Qualifiers
 FH Misc-difference 123
 FT /note= "Encoded by GCC"

XX US5635617-A.

XX 03-JUN-1997.

XX 26-APR-1994; 94US-00233788.

XX 26-APR-1993; 93US-00054452.

XX (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.

XX Collinson SK, Kay WW, Doran JL;

XX WPI; 1997-309886/28.

XX N-PSDB; AAT74142.

XX Isolated Salmonella gene agfa - used for diagnosis of Salmonella or
 PT enteropathogenic bacteria of the Enterobacteria family.

XX Example 2; Fig 7; 85pp; English.

XX The present sequence represents agfa encoded by the full agfa gene
 CC derived from Salmonella enteritidis 27655-3b. The nucleic acid can be

CC used to provide diagnostic assays for Salmonella and/or enteropathogenic
 CC bacteria of the family Enterobacteria. It can also be used to provide
 CC proteins and antibodies which can be used for assays. The nucleic acid
 CC sequence can be used to provide probes or primers which can specifically
 CC hybridise to nucleic acid molecules from greater than 99% of Salmonella
 CC strains that are pathogenic to warm-blooded animals relative to nucleic
 CC acid molecules from virtually all other microbial organisms. (Updated on
 CC 25-MAR-2003 to correct PF field.)

XX Sequence 151 AA;

Query Match 87.9%; Score 686; DB 2; Length 151;
 Best Local Similarity 90.1%; Pred. No. 2.1e-58;
 Matches 136; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

Qy 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGDPDYQLVTRVVTHEMAHALQ 60
 Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGDPSTLSIYQYGSANAALALQ 60
 Qy 61 SDARKSETTITQSGYNGADYVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGOYGG 120
 Db 61 SDARKSETTITQSGYNGADYVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGOYGG 120
 Qy 121 NNAALVNOTASDSSVMVROVGFGNATANQY 151
 Db 121 NNPALVNOTASDSSVMVROVGFGNATANQY 151

RESULT 5

AA36353
 ID AAB36353 standard; protein; 151 AA.

XX AC AAB36353;

XX 26-FEB-2001 (first entry)

XX Agfa::PT3#8 amino acid sequence SEQ ID NO:26.

XX Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.

XX Salmonella enteritidis.

XX Escherichia coli.

XX Synthetic.

XX WO2000060102-A2.

XX 12-OCT-2000.

XX 05-APR-2000; 2000WO-CA000356.

XX 05-APR-1999; 99US-0127888P.

XX (UYVI-) UNIV VICTORIA.

XX White AP, Doran JL, Collinson SK, Kay WW;

XX WPI; 2000-672631/65.

XX N-PSDB; AAC64629.

XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence

XX which encodes foreign epitope or antigen, expresses recombinant Agfa

XX protein useful for eliciting immune response in animal.

XX Disclosure; Page 138; 139pp; English.

XX The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)

CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fibrillar presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fibrillar subunit proteins are usually strong
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

XX Sequence 151 AA;

Query Match 86.5%; Score 675; DB 3; Length 151;
 Best Local Similarity 81.9%; Pred. No. 2.4e-57;
 Matches 136; Conservative 0; Mismatches 0; Indels 30; Gaps 2;
 QY 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGHNGGNSGPP-----YDQ 45
 DB 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGHNGGNSGPPDSTLSIYQYGSANAALQ 60
 QY 46 LVTRVVTHEMAHALQSDARKSETTITQSGYNGADVGOGADNSTIELTQNGFRNNATIDQ 105
 DB 61 LVTRVVTHEMAHA-----GYNGADVGOGADNSTIELTQNGFRNNATIDQ 105
 QY 106 WNAKNSDITVQYGGNNAALVNQTASDSSVMVRQVGFNNATANQY 151
 DB 106 WNAKNSDITVQYGGNNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 6

AAB36350
 ID AAB36350 standard; protein; 151 AA.
 AC AAB36350;
 XX
 XX 26-FEB-2001 (first entry)
 XX Agfa::PT3#5 amino acid sequence SEQ ID NO:20.
 DE
 XX Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX WC200060102-A2.
 XX
 XX 12-OCT-2000.
 PD
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 XX White AP, Doran JL, Collison SK, Kay WW;
 PI
 XX WPI; 2000-672631/65.
 DR N-PSDB; AAC64626.
 XX
 XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PF which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT

PT protein useful for eliciting immune response in animal.
 XX Disclosure; Page 137; 139pp; English.

XX The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/FAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC copy of that gene; and (4) eliciting an immune response in an animal, Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fibrillar presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fibrillar subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

XX Sequence 151 AA;

Query Match 79.7%; Score 622; DB 3; Length 151;
 Best Local Similarity 73.6%; Pred. No. 3.3e-52;
 Matches 128; Conservative 0; Mismatches 0; Indels 46; Gaps 2;
 QY 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGHNGGNSGPP----- 42
 DB 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGHNGGNSGPPDSTLSIYQYGSANAALQ 60
 QY 43 -----YDQLVTRVVTHEMAHALQSDARKSETTITQSGYNGADVGOGADNSTIELTQNGF 97
 DB 61 SDARKYDQLVTRVVTHEMAHA-----GOGADNSTIELTQNGF 97
 QY 98 RNNATIDQWNAKNSDITVQYGGNNAALVNQTASDSSVMVRQVGFNNATANQY 151
 DB 98 RNNATIDQWNAKNSDITVQYGGNNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 7

AAB36348
 ID AAB36348 standard; protein; 151 AA.
 AC AAB36348;
 XX
 XX 26-FEB-2001 (first entry)
 XX Agfa::PT3#3 amino acid sequence SEQ ID NO:16.
 DE
 XX Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX WC200060102-A2.
 XX
 XX 12-OCT-2000.
 PD
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 XX

```
PR 05-APR-1999; 99US-0127888P.
XX (UYVI-) UNIV VICTORIA.
XX White AP, Doran JL, Collison SK, Kay WW;
XX WPI; 2000-672631/65.
DR N-PSDB; AAC64624.
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 136; 139pp; English.
XX
XX The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
XX Sequence 151 AA;
XX
XX Query Match 79.2%; Score 618; DB 3; Length 151;
XX Best Local Similarity 76.6%; Pred. No. 8e-52;
XX Matches 131; Conservative 0; Mismatches 0; Indels 40; Gaps 2;
XX
XX QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDYDQLVTRVVTHEMAH--- 57
XX DB 1 MKLLKVAFAAIVVSGSALAGV-----YDQLVTRVVTHEMAHAGS 40
XX
XX QY 58 -----ALQSDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNN 100
XX DB 41 PDSTLSIYQYGSANAALQSDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNN 100
XX
XX QY 101 ATIDOWNAKNSDITVQYGGNNAALVNOTASDSSVWVQVGFNNATANQY 151
XX DB 101 ATIDOWNAKNSDITVQYGGNNAALVNOTASDSSVWVQVGFNNATANQY 151
XX
XX RESULT 8
XX ID AAB36346
XX
XX XX AAB36346 standard; protein; 151 AA.
XX
XX AC AAB36346;
XX
XX DT 26-FEB-2001 (first entry)
XX
XX XX Agfa::PT3#1 amino acid sequence SEQ ID NO:12.
XX
XX KW Salmonella; agfa; chromosomal gene replacement; fimbrial; epitope;
XX vaccine; immune response; immunogen.
```

```
XX Salmonella enteritidis.
XX Escherichia coli.
XX Synthetic.
XX WO200060102-A2.
XX 12-OCT-2000.
XX
XX 05-APR-2000; 2000WO-CA000356.
XX
XX 05-APR-1999; 99US-0127888P.
XX (UYVI-) UNIV VICTORIA.
XX
XX White AP, Doran JL, Collison SK, Kay WW;
XX WPI; 2000-672631/65.
DR N-PSDB; AAC64622.
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 135; 139pp; English.
XX
XX The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
XX Sequence 151 AA;
XX
XX Query Match 78.6%; Score 613; DB 3; Length 151;
XX Best Local Similarity 80.8%; Pred. No. 2.4e-51;
XX Matches 122; Conservative 7; Mismatches 22; Indels 0; Gaps 0;
XX
XX QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDYDQLVTRVVTHEMAHALQ 60
XX DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
XX
XX QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITVQYGG 120
XX DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITVQYGG 120
XX
XX QY 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151
XX DB 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151
XX
XX RESULT 9
```

AAB36347
 ID AAB36347 standard; protein; 151 AA.
 XX
 AC AAB36347;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Agfa::PT3#2 amino acid sequence SEQ ID NO:14.
 XX
 KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
 XX vaccine; immune response; immunogen.
 OS Salmorella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 PN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay WW;
 XX
 DR WPI: 2000-672631/65.
 DR N-PSDB; AAC64623.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 136; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEFI7/TAFF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC the exemplification of the present invention

Query Match 78.3%; Score 611; DB 3; Length 151;
 Best Local Similarity 81.5%; Pred. NO. 3.8e-51;
 Matches 123; Conservative 6; Mismatches 22; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGGNSGGPDYDQLVTRVVVTHENAHLQ 60
 DB 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGGNSGGPDYDQLVTRVVVTHENAHLQ 60

61 SDARKSETTITQSGYNGADVQGCADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGG 120
 |||||
 61 SDARKSETTITQSGYNGADVQGCADNSTIELTQNGFRNNATIDQWNAKNSDITVQYDQ 120
 |||||
 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151
 : : : : :
 121 LVTRVVTHENAHASVMVRQVGFNNATANQY 151
 : : : : :
 RESULT 10
 AAB36352
 ID AAB36352 standard; protein; 151 AA.
 XX
 AC AAB36352;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Agfa::PT3#7 amino acid sequence SEQ ID NO:24.
 XX
 KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
 XX vaccine; immune response; immunogen.
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 PN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay WW;
 XX
 DR WPI: 2000-672631/65.
 DR N-PSDB; AAC64628.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 138; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEFI7/TAFF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC the exemplification of the present invention

CC the exemplification of the present invention
 XX Sequence 151 AA;
 SQ

Query Match 77.1%; Score 608; DB 3; Length 151;
 Best Local Similarity 82.1%; Pred. No. 7.4e-51;
 Matches 124; Conservative 5; Mismatches 22; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDYDQLVTRVVTHEMAHALQ 60
 DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQVGSANAALQ 60
 QY 61 SDARKSETTITQSGYNGADYVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGOYGG 120
 DB 61 SDARKSETTITQSGYNGADYVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVTRVVT 120

QY 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151
 DB 121 HEMAHANQASDSSVMVRQVGFNNATANQY 151

RESULT 11
 AAB36354
 ID AAB36354 standard; protein; 151 AA.
 AC AAB36354;
 DT 26-FEB-2001 (first entry)
 DE AgfA::PT3#9 amino acid sequence SEQ ID NO:28.
 KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
 KW vaccine; immune response; immunogen.
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 PN WO200050102-A2.
 XX 12-OCT-2000.
 XX 05-APR-2000; 2000WO-CA000356.
 XX 05-APR-1999; 99US-0127888P.
 PA (UYVI-) UNIV VICTORIA.
 PI White AP, Doran JL, Collison SK, Kay WW;
 XX WPI; 2000-672631/65.
 DR N-PSDB; AAC64630.
 CC Recombinant agfA gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant AgfA
 PT protein useful for eliciting immune response in animal.
 XX Disclosure; Page 138; 139pp; English.

The present invention describes a recombinant agfA gene (I) where a
 segment of the gene has been replaced by a segment of a foreign DNA
 sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC AgfA, CsgA and AgfA-homologue fimbria subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant AgfA
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or

CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine. The carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX Sequence 151 AA;
 SQ

Query Match 77.1%; Score 601; DB 3; Length 151;
 Best Local Similarity 80.8%; Pred. No. 3.5e-50;
 Matches 122; Conservative 4; Mismatches 25; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDYDQLVTRVVTHEMAHALQ 60
 DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQVGSANAALQ 60
 QY 61 SDARKSETTITQSGYNGADYVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGOYGG 120
 DB 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFRNNATIDQWNAKNSDITVGOYGG 120
 QY 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151
 DB 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151

RESULT 12
 AAB36351
 ID AAB36351 standard; protein; 151 AA.
 AC AAB36351;
 DT 26-FEB-2001 (first entry)
 DE AgfA::PT3#6 amino acid sequence SEQ ID NO:22.
 KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
 KW vaccine; immune response; immunogen.
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 PN WO200060102-A2.
 XX 12-OCT-2000.
 XX 05-APR-2000; 2000WO-CA000356.
 XX 05-APR-1999; 99US-0127888P.
 PA (UYVI-) UNIV VICTORIA.
 PI White AP, Doran JL, Collison SK, Kay WW;
 XX WPI; 2000-672631/65.
 DR N-PSDB; AAC64627.
 CC Recombinant agfA gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant AgfA
 PT protein useful for eliciting immune response in animal.
 XX Disclosure; Page 137; 139pp; English.

The present invention describes a recombinant agfA gene (I) where a
 segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:

CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA, and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX Sequence 151 AA;
 SQ

Query Match 76.9%; Score 600; DB 3; Length 151;
 Best Local Similarity 81.5%; Pred. No. 4.4e-50;
 Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGSGGPDYDQLVTRVVTHEMAHALQ 60
 DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGSGGPDSTLSIYQGSANAALALQ 60

QY 61 SPARKSETTITQSGYNGADVGQGDADNSTIELTQNGFRNATIDQWNAKNSDITVGYGG 120
 DB 61 SPARKSETTITQSGYNGADVGQGDADNSTIELTQNGFRNATIDQWNAKNSDITVGYGG 120

QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
 DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 13
 AAB36355
 ID AAB36355 standard; protein; 151 AA.
 AC AAB36355;
 XX
 XX 26-FEB-2001 (first entry)
 DE Agfa::PT3#10 amino acid sequence SEQ ID NO:30.
 KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX WO2000060102-A2.
 XX 12-OCT-2000.
 XX 05-APR-2000; 2000WO-CA000356.
 XX 05-APR-1999; 99US-0127888P.
 XX (UYVI-) UNIV VICTORIA.
 XX White AP, Doran JL, Collison SK, Kay WW;
 PI WPI; 2000-672631/65.
 DR

DR N-PSDB; AAC64631.
 XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 139; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA, and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX Sequence 151 AA;
 SQ

Query Match 76.8%; Score 599; DB 3; Length 151;
 Best Local Similarity 81.5%; Pred. No. 5.5e-50;
 Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGSGGPDYDQLVTRVVTHEMAHALQ 60
 DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGSGGPDSTLSIYQGSANAALALQ 60

QY 61 SPARKSETTITQSGYNGADVGQGDADNSTIELTQNGFRNATIDQWNAKNSDITVGYGG 120
 DB 61 SPARKSETTITQSGYNGADVGQGDADNSTIELTQNGFRNATIDQWNAKNSDITVGYGG 120

QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
 DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 14
 AAB36343
 ID AAB36343 standard; protein; 151 AA.
 AC AAB36343;
 XX
 XX 26-FEB-2001 (first entry)
 DE Escherichia coli CsgA amino acid sequence SEQ ID NO:7.
 KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 OS Escherichia coli.
 XX WO2000060102-A2.
 XX 12-OCT-2000.
 DR

PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay WW;
 XX
 DR WPI; 2000-672631/65.
 DR N-PSDB; AAC64619.
 XX
 XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant AgfA
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 135; 139pp; English.
 XX
 CC The present invention describes a recombinant agfA gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (Sf17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC AgfA, CsgA and AgfA-homologue fibrin subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant AgfA
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fibrillar presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fibrin protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fibrillar subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;
 Query Match 67.1%; Score 523; DB 3; Length 151;
 Best Local Similarity 68.9%; Pred. No. 1.2e-42; Mismatches 27; Indels 0; Gaps 0;
 Matches 104; Conservative 20; Mismatches 27; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHGGNGSSGPDYDQLVTRVVTHEMAHALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHGGNGSSGPDYDQLVTRVVTHEMAHALQ 60
 QY 61 SDARKSETITQSGYGNAGADVGGADNSTIELTQNGFRNNATIDOWNAKNSDITVQYGG 120
 DB 61 TDARNSDLTITQGGGNGADVGGSDSSIDLITQRFNGSATLDQWNGKNSMTVKQFGG 120
 QY 121 NNAALVNOTASDSSVVMVQVGFNNATANQY 151
 DB 121 GNGAAVDQTASNSVNVTVQVGFNNATAHQY 151
 RESULT 15
 ID ABR82651
 XX ABR82651 standard; protein; 151 AA.
 AC ABR82651;
 XX
 DT 04-DEC-2003 (first entry)
 XX
 DE E. coli CsgA subunit 15 kDa protein.
 XX

Plasma protein; immune response; antibacterial; vaccine; gene therapy.
 Escherichia coli.
 WO2003064446-A2.
 07-AUG-2003.
 30-JAN-2003; 2003WO-EP000943.
 31-JAN-2002; 2002GB-00002275.
 (HANS-) HANSA MEDICAL RES AB.
 Bjoerck L, Olsen A, Wikstroem M, Herwald H;
 WPI; 2003-646136/61.
 N-PSDB; ACF36153.
 New isolated peptide capable of binding a mammalian plasma protein,
 useful in the manufacture of a medicament for the prevention and/or
 treatment of a bacterial infection, such as Escherichia coli, Salmonella
 or Shigella infections.
 Disclosure; Page 41-42; 42pp; English.
 The invention relates to an isolated peptide capable of binding a
 mammalian plasma protein or of generating an immune response in a mammal
 selected from sequences shown in ABR82642, ABR82648-49. The peptide or
 antibody is useful for treating a bacterial infection in a human or
 animal or in the manufacture of a medicament for the prophylactic
 treatment of a bacterial infection, such as Escherichia coli, Salmonella
 or Shigella infection. The peptide that is immobilized on a solid support
 is also useful as a reagent for determining the ability of a plasma
 protein to bind to bacteria. The present sequence represents an E. coli
 15 kDa protein
 Sequence 151 AA;
 Query Match 66.4%; Score 518; DB 7; Length 151;
 Best Local Similarity 68.2%; Pred. No. 3.8e-42;
 Matches 103; Conservative 20; Mismatches 28; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHGGNGSSGPDYDQLVTRVVTHEMAHALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHGGNGSSGPDYDQLVTRVVTHEMAHALQ 60
 QY 61 SDARKSETITQSGYGNAGADVGGADNSTIELTQNGFRNNATIDOWNAKNSDITVQYGG 120
 DB 61 TDARNSDLTITQGGGNGADVGGSDSSIDLITQRFNGSATLDQWNGKNSMTVKQFGG 120
 QY 121 NNAALVNOTASDSSVVMVQVGFNNATANQY 151
 DB 121 GNGAAVDQTASNSVNVTVQVGFNNATAHQY 151
 Search completed: August 2, 2004, 14:48:25
 Job time : 44.9 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:40:48 ; Search time 12 Seconds
(without alignments)

649.627 Million cell updates/sec

Title: US-09-543-407-18

Perfect score: 780

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVQVGFNNATANYQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	686	87.9	151	1	US-08-233-788A-59
2	506	64.9	120	1	US-08-233-788A-57
3	92	11.8	975	4	US-09-328-352-4764
4	90	11.5	906	1	US-08-254-573-2
5	90	11.5	906	1	US-08-687-379-2
6	90	11.5	906	1	US-08-687-379-4
7	90	11.5	906	4	US-08-172-332-1
8	90	11.5	906	4	US-08-216-348-2
9	87.5	11.2	738	3	US-08-864-038A-3
10	84	10.8	892	4	US-09-336-447A-5
11	84	10.8	907	1	US-07-718-575-2
12	84	10.8	907	1	US-08-481-206-2
13	84	10.8	907	2	US-08-486-269A-2
14	84	10.8	943	4	US-09-056-586-204
15	84	10.8	943	4	US-09-072-596-199
16	84	10.8	943	4	US-09-477-135A-131
17	84	10.8	943	4	US-09-072-967-204
18	84	10.8	1415	4	US-09-253-991A-26438
19	83.5	10.7	415	4	US-09-025-769B-280
20	83.5	10.7	873	4	US-09-336-447A-13
21	83	10.6	518	3	US-09-043-123-2
22	83	10.6	1912	1	US-08-409-995-4
23	83	10.6	1912	3	US-08-685-467-4
24	83	10.6	2353	3	US-09-377-155-33
25	83	10.6	2353	3	US-08-913-942-4
26	83	10.6	2353	4	US-09-669-974-33
27	83	10.6	2353	4	US-09-797-862-33

ALIGNMENTS

RESULT 1

US-08-233-788A-59
; Sequence 59, Application US/08233788A

; Patent No. 5635617

; GENERAL INFORMATION:

; APPLICANT: Doran, James L.

; APPLICANT: Kay, William W.

; APPLICANT: Collinson, Karen S.

; APPLICANT: Clouthier, Sharon C.

; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION

; TITLE OF INVENTION: OF SALMONELLA

; NUMBER OF SEQUENCES: 61

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Seed and Berry

; STREET: 6300 Columbia Center, 701 Fifth Avenue

; CITY: Seattle

; STATE: Washington

; COUNTRY: U.S.A.

; ZIP: 98104-7092

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/233,788A

; FILING DATE: 26-APR-1994

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: King, Joshua

; REGISTRATION NUMBER: 35,570

; REFERENCE/DOCKET NUMBER: 920043.403C2

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (206) 622-4900

; TELEFAX: (206) 682-6031

; TELEX: 3723836 SEEDANBERRY

; INFORMATION FOR SEQ ID NO: 59:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 151 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-08-233-788A-59

Query Match 87.9%; Score 686; DB 1; Length 151;

Best Local Similarity 90.1%; Pred. No. 2.2e-62;

Matches 136; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

Qy 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNNHNGGNSGPDYDQLVTRVVTTHMAHALQ 60

|||||

Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSSGPDSTLSIYQYGSANAALQ 60
QY 61 SDARKSETTITQSGYNGADVGCGADNSTIETQNGFRNNAIDQWNAKNSDITVGYGG 120
Db 61 SDARKSETTITQSGYNGADVGCGADNSTIETQNGFRNNAIDQWNAKNSDITVGYGG 120
QY 121 NNAALVNQASDSSVMVRQVGFNNATANQY 151
Db 121 NNPALVNQASDSSVMVRQVGFNNATANQY 151

RESULT 2

US-08-233-788A-57
; Sequence 57, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, Karen S.
; APPLICANT: Clouthier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,788A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 120 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-233-788A-57

Query Match 64.9%; Score 506; DB 1; Length 120;
Best Local Similarity 87.5%; Pred. No. 3, 4e-44;
Matches 98; Conservative 3; Mismatches 11; Indels 0; Gaps 0;
QY 22 VVPQWGGGNGHNGSSGPDYDQVLTIVRVVTHMAHALQSDARKSETTITQSGYNGADY 81
Db 1 VVPQWGGGNGHNGSSGPDSTLSIYQYGSANAALQSDARKSETTITQSGYNGADY 60
QY 82 GCGADNSTIETQNGFRNNAIDQWNAKNSDITVGYGNNALVNQASPS 133
Db 61 GCGADNSTIETQNGFRNNAIDQWNAKNSDITVGYGNNALVNQASPS 112

RESULT 3

US-09-328-352-4764
; Sequence 4764, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:

; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4764
; LENGTH: 975
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-4764

Query Match 11.8%; Score 92; DB 4; Length 975;
Best Local Similarity 23.8%; Pred. No. 0.75;
Matches 35; Conservative 24; Mismatches 52; Indels 36; Gaps 7;
QY 15 SGSALAGVVPQWGGGNGHNGG-GNSGPDYDQVLTIVRVVTHM-----AHALQSDA 63
Db 300 AGNGIA-----SCNGEHNYGIGNGDDVD--ITAPITGVNLISGNSFTLIGNSSSV 351
QY 64 RKSETTITQS-----GYNGADVGCGADNSTIETQNGF-----RNNATIDQWNAKNS 111
Db 352 NTAPTTSTNVNDNTIDNGSGTSGSGNGSGDGLLNGASNGEHNYGIGNGDDVD 411
QY 112 DIT-----VGQYGGNNAALVNQASPS 133
Db 412 DITSPITGTFNFGNSGSLIGNSSSS 438

RESULT 4

US-08-254-573-2
; Sequence 2, Application US/08254573
; Patent No. 5610032
; GENERAL INFORMATION:
; APPLICANT: KAMBOJ, Rajender
; APPLICANT: ELLIOTT, Candace
; APPLICANT: NUTT, Stephen
; TITLE OF INVENTION: AMPA-BINDING HUMAN GLUR1 RECEPTORS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/254,573
; FILING DATE: 06-JUN-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/896,611
; FILING DATE: 10-JUN-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 16777/179 ALLE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-9300
; TELEFAX: (703) 883-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 906 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

US-08-254-573-2
Query Match 11.5%; Score 90; DB 1; Length 906;
Best Local Similarity 22.6%; Pred. No. 1.1;
Matches 45; Conservative 26; Mismatches 56; Indels 72; Gaps 10;
QY 6 VAAFAAIVVSGSALAGVVPQWGGGNGGNSGGPDYDQL-----VTRVVTHE-----M 55
DB 250 VTGFQLVNVTDTIPAKIMQW-----KNSDARDHTRVDWKRPKYTSALTYDGKVM 300
QY 56 AHALQSDARKSETTITQSGYNGADV-----GOGAD-----NSTIELTQ 94
DB 301 AEAFOQLRRQ---RIDISRRGNAGDCLANPAVFWGGGIDIQALQQRVEGLTGNVQFNE 357
QY 95 NGRFRNAT-----IDQWNAKNSDI---TVQYGGNNAALVNQT-----ASD 132
DB 358 KGRRTNYTLHVIMKHDGIRKIGYWNEDDKFVPAATDAQAGDSSVQNRITVITTTILED 417
QY 133 SSVWVRQVGFNNATANQY 151
DB 418 PYVMLKK-----NANQF 429
RESULT 5
US-08-687-379-2
; Sequence 2, Application US/08687379
; Patent No. 5756897
; GENERAL INFORMATION:
; APPLICANT: Hoeger, Thomas
; APPLICANT: Utsch, Andreas
; APPLICANT: Bach, Alfred
; APPLICANT: Sterrer, Sylvia
; APPLICANT: Lemaire, Hans-Georg
; TITLE OF INVENTION: Subunits of Glutamate Receptors, Their
; PREPARATION AND THEIR USE
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Keil & Weinkauff
; STREET: 1101 Connecticut Avenue
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette-3.5 inch, 720 Kb storage
; COMPUTER: IBM AT-compatible, 80286 processor
; OPERATING SYSTEM: MS-DOS version 6.0
; SOFTWARE: WordPerfect version 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/687,379
; FILING DATE: 05-AUG-1996
; CLASSIFICATION: 435
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 906 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-687-379-2
Query Match 11.5%; Score 90; DB 1; Length 906;
Best Local Similarity 22.6%; Pred. No. 1.1;
Matches 45; Conservative 26; Mismatches 56; Indels 72; Gaps 10;
QY 6 VAAFAAIVVSGSALAGVVPQWGGGNGGNSGGPDYDQL-----VTRVVTHE-----M 55
DB 250 VTGFQLVNVTDTIPAKIMQW-----KNSDARDHTRVDWKRPKYTSALTYDGKVM 300
QY 56 AHALQSDARKSETTITQSGYNGADV-----GOGAD-----NSTIELTQ 94
DB 301 AEAFOQLRRQ---RIDISRRGNAGDCLANPAVFWGGGIDIQALQQRVEGLTGNVQFNE 357
QY 95 NGRFRNAT-----IDQWNAKNSDI---TVQYGGNNAALVNQT-----ASD 132
DB 358 KGRRTNYTLHVIMKHDGIRKIGYWNEDDKFVPAATDAQAGDSSVQNRITVITTTILED 417
QY 133 SSVWVRQVGFNNATANQY 151
DB 418 PYVMLKK-----NANQF 429
RESULT 6
US-08-687-379-4
; Sequence 4, Application US/08687379
; Patent No. 5756697
; GENERAL INFORMATION:
; APPLICANT: Hoeger, Thomas
; APPLICANT: Utsch, Andreas
; APPLICANT: Bach, Alfred
; APPLICANT: Sterrer, Sylvia
; APPLICANT: Lemaire, Hans-Georg
; TITLE OF INVENTION: Subunits of Glutamate Receptors, Their
; PREPARATION AND THEIR USE
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Keil & Weinkauff
; STREET: 1101 Connecticut Avenue
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette-3.5 inch, 720 Kb storage
; COMPUTER: IBM AT-compatible, 80286 processor
; OPERATING SYSTEM: MS-DOS version 6.0
; SOFTWARE: WordPerfect version 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/687,379
; FILING DATE: 05-AUG-1996
; CLASSIFICATION: 435
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 906 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-687-379-4
Query Match 11.5%; Score 90; DB 1; Length 906;
Best Local Similarity 22.6%; Pred. No. 1.1;
Matches 45; Conservative 26; Mismatches 56; Indels 72; Gaps 10;
QY 6 VAAFAAIVVSGSALAGVVPQWGGGNGGNSGGPDYDQL-----VTRVVTHE-----M 55
DB 250 VTGFQLVNVTDTIPAKIMQW-----KNSDARDHTRVDWKRPKYTSALTYDGKVM 300
QY 56 AHALQSDARKSETTITQSGYNGADV-----GOGAD-----NSTIELTQ 94
DB 301 AEAFOQLRRQ---RIDISRRGNAGDCLANPAVFWGGGIDIQALQQRVEGLTGNVQFNE 357
QY 95 NGRFRNAT-----IDQWNAKNSDI---TVQYGGNNAALVNQT-----ASD 132
DB 358 KGRRTNYTLHVIMKHDGIRKIGYWNEDDKFVPAATDAQAGDSSVQNRITVITTTILED 417
QY 133 SSVWVRQVGFNNATANQY 151
DB 418 PYVMLKK-----NANQF 429
RESULT 7
US-08-172-332-1
; Sequence 1, Application US/08172332
; Patent No. 6313279
; GENERAL INFORMATION:
; APPLICANT: Burnett, J. Paul
; APPLICANT: Mayne, Nancy G

US-08-254-573-2
Query Match 11.5%; Score 90; DB 1; Length 906;
Best Local Similarity 22.6%; Pred. No. 1.1;
Matches 45; Conservative 26; Mismatches 56; Indels 72; Gaps 10;
QY 6 VAAFAAIVVSGSALAGVVPQWGGGNGGNSGGPDYDQL-----VTRVVTHE-----M 55
DB 250 VTGFQLVNVTDTIPAKIMQW-----KNSDARDHTRVDWKRPKYTSALTYDGKVM 300
QY 56 AHALQSDARKSETTITQSGYNGADV-----GOGAD-----NSTIELTQ 94
DB 301 AEAFOQLRRQ---RIDISRRGNAGDCLANPAVFWGGGIDIQALQQRVEGLTGNVQFNE 357
QY 95 NGRFRNAT-----IDQWNAKNSDI---TVQYGGNNAALVNQT-----ASD 132
DB 358 KGRRTNYTLHVIMKHDGIRKIGYWNEDDKFVPAATDAQAGDSSVQNRITVITTTILED 417
QY 133 SSVWVRQVGFNNATANQY 151
DB 418 PYVMLKK-----NANQF 429
RESULT 5
US-08-687-379-2
; Sequence 2, Application US/08687379
; Patent No. 5756897
; GENERAL INFORMATION:
; APPLICANT: Hoeger, Thomas
; APPLICANT: Utsch, Andreas
; APPLICANT: Bach, Alfred
; APPLICANT: Sterrer, Sylvia
; APPLICANT: Lemaire, Hans-Georg
; TITLE OF INVENTION: Subunits of Glutamate Receptors, Their
; PREPARATION AND THEIR USE
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Keil & Weinkauff
; STREET: 1101 Connecticut Avenue
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette-3.5 inch, 720 Kb storage
; COMPUTER: IBM AT-compatible, 80286 processor
; OPERATING SYSTEM: MS-DOS version 6.0
; SOFTWARE: WordPerfect version 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/687,379
; FILING DATE: 05-AUG-1996
; CLASSIFICATION: 435
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 906 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-687-379-2
Query Match 11.5%; Score 90; DB 1; Length 906;
Best Local Similarity 22.6%; Pred. No. 1.1;
Matches 45; Conservative 26; Mismatches 56; Indels 72; Gaps 10;
QY 6 VAAFAAIVVSGSALAGVVPQWGGGNGGNSGGPDYDQL-----VTRVVTHE-----M 55
DB 250 VTGFQLVNVTDTIPAKIMQW-----KNSDARDHTRVDWKRPKYTSALTYDGKVM 300
QY 56 AHALQSDARKSETTITQSGYNGADV-----GOGAD-----NSTIELTQ 94
DB 301 AEAFOQLRRQ---RIDISRRGNAGDCLANPAVFWGGGIDIQALQQRVEGLTGNVQFNE 357
QY 95 NGRFRNAT-----IDQWNAKNSDI---TVQYGGNNAALVNQT-----ASD 132

APPLICANT: Snyder, Yvonne M
APPLICANT: Sharp, Robert L
TITLE OF INVENTION: HUMAN GLUTAMATE RECEPTOR AND RELATED DNA
TITLE OF INVENTION: COMPOUNDS
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lilly Patent Division/JPL
STREET: Lilly Corporate Center
CITY: Indianapolis
STATE: IN
COUNTRY: USA
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/172,332
FILING DATE: 22-DEC-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/879,688
FILING DATE: May 1, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Leeds, James P.
REGISTRATION NUMBER: 35241
REFERENCE/DOCKET NUMBER: X-8342
TELEPHONE: 317-276-1667
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 906 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-172-332-1

Query Match 11.5%; Score 90; DB 4; Length 906;
Best Local Similarity 22.6%; Pred. No. 1.1;
Matches 45; Conservative 26; Mismatches 56; Indels 72; Gaps 10;

QY 6 VAAFAAIVVSGSALAGVWPQWGGGNGHNGSGSPDYDQL-----VTRVVTHE---M 55
DB 250 VTGFQLVNYDTIPAKIMQW-----KNSDARDHTRVDWKRPKYSALTIDGVKVM 300

QY 56 AHALQSDARKSETTTTQSGYNGADV-----GQGD-----NSTIELTQ 94
DB 301 AEAFOQLRRQ---RIDISRRGNAGDCLANPAVPWQGGIDIQALQVREFGLTGNVQFNE 357
DB 95 NGRNNAT-----IDQNAKNSDI---TVQYGGNNAALVNQT-----ASD 132
DB 358 KGRRTNYTHLVEMKHDGIRKGYWNEDDKFVPAATDAQAGDSSVQNRITYVTITLED 417

QY 133 SSVVVRQVGFNGNATANQY 151
DB 418 PYVMLKK-----NANQF 429

RESULT 8
US-08-216-326-2
Sequence 2, Application US/08216326
Patent No. 6406868
GENERAL INFORMATION:
APPLICANT: KAMBOJ, Rajender
APPLICANT: ELIOTT, Candace
APPLICANT: NUTT, Stephen
TITLE OF INVENTION: AMPA-BINDING HUMAN GLUR1 RECEPTORS
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street N.W., Suite 500
CITY: Washington, D.C.

COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/216,326
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/896,611
FILING DATE: 10-JUN-1992
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 16777/229/ALLE
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 906 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-216-326-2

Query Match 11.5%; Score 90; DB 4; Length 906;
Best Local Similarity 22.6%; Pred. No. 1.1;
Matches 45; Conservative 26; Mismatches 56; Indels 72; Gaps 10;

QY 6 VAAFAAIVVSGSALAGVWPQWGGGNGHNGSGSPDYDQL-----VTRVVTHE---M 55
DB 250 VTGFQLVNYDTIPAKIMQW-----KNSDARDHTRVDWKRPKYSALTIDGVKVM 300

QY 56 AHALQSDARKSETTTTQSGYNGADV-----GQGD-----NSTIELTQ 94
DB 301 AEAFOQLRRQ---RIDISRRGNAGDCLANPAVPWQGGIDIQALQVREFGLTGNVQFNE 357
QY 95 NGRNNAT-----IDQNAKNSDI---TVQYGGNNAALVNQT-----ASD 132
DB 358 KGRRTNYTHLVEMKHDGIRKGYWNEDDKFVPAATDAQAGDSSVQNRITYVTITLED 417

QY 133 SSVVVRQVGFNGNATANQY 151
DB 418 PYVMLKK-----NANQF 429

RESULT 9
US-08-864-038A-3
Sequence 3, Application US/08864038A
Patent No. 6001592
GENERAL INFORMATION:
APPLICANT: Kunio NAKASHIMA et al.
TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE cDNA, VECTOR
TITLE OF INVENTION: CONTAINING SAID cDNA, HOST CELLS TRANSFORMED WITH SAID
TITLE OF INVENTION: VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING
TITLE OF INVENTION: SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE
TITLE OF INVENTION: TO SAID POLYPEPTIDE
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: 812-5 Hirano
STREET: Ishinden
CITY: Tsu-city
STATE: Mie-prefecture
COUNTRY: JAPAN
ZIP: 514-01
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: Microsoft Windows 95

QY 142 FGNNATANOY 151
 Db 425 -----NANOF 429

RESULT 12
 US-08-481-206-2
 ; Sequence 2, Application US/08481206
 ; Patent No. 5739291
 ; GENERAL INFORMATION:
 ; APPLICANT: Heinemann Ph.D., Stephen P.
 ; APPLICANT: Boulter Ph.D., James R.
 ; APPLICANT: Hollmann Ph.D., Michael NMN
 ; APPLICANT: Bettler Ph.D., Bernhard NMN
 ; APPLICANT: Jensen Ph.D., Jan E.
 ; TITLE OF INVENTION: GLUTAMATE RECEPTOR COMPOSITIONS AND
 ; TITLE OF INVENTION: METHODS
 ; NUMBER OF SEQUENCES: 14
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
 ; STREET: 444 So. Flower St., Suite 2000
 ; CITY: Los Angeles
 ; STATE: California
 ; COUNTRY: United States
 ; ZIP: 90071-2921
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/481,206
 ; FILING DATE:
 ; CLASSIFICATION: 530
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/013,767
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Reiter Ph.D., Stephen E.
 ; REGISTRATION NUMBER: 31192
 ; REFERENCE/DOCKET NUMBER: P31 8962
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (619) 535-9001
 ; TELEFAX: (619) 535-8949
 ; TELEX: 9103330318
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 907 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-481-206-2

Query Match 10.8%; Score 84; DB 1; Length 907;
 Best Local Similarity 22.6%; Pred. No. 4.4;
 Matches 43; Conservative 24; Mismatches 69; Indels 54; Gaps 8;
 QY 6 VAAFAIVVSGSALAGVPPQWGGGNNHGGG-NSSGPDYDQLVTRVVTHEMAHALQSDAR 64
 Db 250 VTGFQLVNYTDTIPARIMQOWRTSDSRDTRVDWKRPKYTSALTVDGVKVMAEAFQSLRR 309
 QY 65 KSETTTTQSGYNGADV-----GQAD-----NSTIELTQNGPRNNAT- 102
 Db 310 Q---RIDISRRNAGDCLANPAVPWGQGDIDIALQOQVRFEGLTGNVQFNEKGRNTYTL 366
 QY 103 -----IDOWNAKNSDI---TVQGYGNNAAALVNQT-----ASDSSVMVRQVG 141
 Db 367 HVIEMKHGDGIRKIGYWNEDDKFVPAATDAQAGDGNSSVQNRITYIVTTILEDPFYVMLKK-- 424
 QY 142 FGNNATANOY 151
 Db 425 -----NANOF 429

RESULT 13
 US-08-486-269A-2
 ; Sequence 2, Application US/08486269A
 ; Patent No. 5945509
 ; GENERAL INFORMATION:
 ; APPLICANT: Heinemann, Stephen P.
 ; APPLICANT: Boulter, James R.
 ; APPLICANT: Hollmann, Michael
 ; APPLICANT: Bettler, Bernhard
 ; APPLICANT: Jensen, Jan E.
 ; TITLE OF INVENTION: GLUTAMATE RECEPTOR COMPOSITIONS
 ; TITLE OF INVENTION: AND METHODS
 ; NUMBER OF SEQUENCES: 14
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Gray Cary Ware & Freidenrich LLP
 ; STREET: 4365 Executive Drive, Suite 1600
 ; CITY: San Diego
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 92121
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FastSeq for Windows DEMONSTRATION Version 2.0D
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/486,269A
 ; FILING DATE: 06-JUN-1995
 ; CLASSIFICATION: 530
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/013,767
 ; FILING DATE: 04-FEB-1993
 ; APPLICATION NUMBER: 07/718,575
 ; FILING DATE: 21-JUN-1991
 ; APPLICATION NUMBER: PCT/US90/06153
 ; FILING DATE: 25-OCT-1990
 ; APPLICATION NUMBER: 07/428,116
 ; FILING DATE: 27-OCT-1989
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Reiter, Stephen E.
 ; REGISTRATION NUMBER: 31,192
 ; REFERENCE/DOCKET NUMBER: P41 9986
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 619-677-1409
 ; TELEFAX: 619-677-1465
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 907 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-486-269A-2

Query Match 10.8%; Score 84; DB 2; Length 907;
 Best Local Similarity 22.6%; Pred. No. 4.4;
 Matches 43; Conservative 24; Mismatches 69; Indels 54; Gaps 8;
 QY 6 VAAFAIVVSGSALAGVPPQWGGGNNHGGG-NSSGPDYDQLVTRVVTHEMAHALQSDAR 64
 Db 250 VTGFQLVNYTDTIPARIMQOWRTSDSRDTRVDWKRPKYTSALTVDGVKVMAEAFQSLRR 309
 QY 65 KSETTTTQSGYNGADV-----GQAD-----NSTIELTQNGPRNNAT- 102
 Db 310 Q---RIDISRRNAGDCLANPAVPWGQGDIDIALQOQVRFEGLTGNVQFNEKGRNTYTL 366
 QY 103 -----IDOWNAKNSDI---TVQGYGNNAAALVNQT-----ASDSSVMVRQVG 141
 Db 367 HVIEMKHGDGIRKIGYWNEDDKFVPAATDAQAGDGNSSVQNRITYIVTTILEDPFYVMLKK-- 424


```
QY 142 FGNATANOY 151
Db 425 -----NANQF 429

RESULT 14
US-09-056-556-204
; Sequence 204, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 204:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 943 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-09-056-556-204
Query Match 10.8%; Score 84; DB 4; Length 943;
Best Local Similarity 25.2%; Pred. No. 4.7;
Matches 31; Conservative 9; Mismatches 41; Indels 42; Gaps 4;

QY 15 SGSALAGVVPQWGGGNGHNGSGSPDYDQLVTRVVTHEMAHALQSDARKSETTITQSG 74
Db 184 SGTGNVGI-----GNSGTGNWIGNSGNSYN-----TG 211

QY 75 YNGADVCGGADNSTIELTQNGFRNATIDQNAKNSD---ITVGOY-----GGNAA 124
Db 212 FGNSGDANTGFNSGIANTGVGNAGNYNTGSPNSNTGTFNMGQYNTGLNSGNYNTG 271

QY 125 LVN 127
Db 272 LAN 274

RESULT 15
US-09-072-596-199
; Sequence 199, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond

; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 199:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 943 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-09-072-596-199
Query Match 10.8%; Score 84; DB 4; Length 943;
Best Local Similarity 25.2%; Pred. No. 4.7;
Matches 31; Conservative 9; Mismatches 41; Indels 42; Gaps 4;

QY 15 SGSALAGVVPQWGGGNGHNGSGSPDYDQLVTRVVTHEMAHALQSDARKSETTITQSG 74
Db 184 SGTGNVGI-----GNSGTGNWIGNSGNSYN-----TG 211

QY 75 YNGADVCGGADNSTIELTQNGFRNATIDQNAKNSD---ITVGOY-----GGNAA 124
Db 212 FGNSGDANTGFNSGIANTGVGNAGNYNTGSPNSNTGTFNMGQYNTGLNSGNYNTG 271

QY 125 LVN 127
Db 272 LAN 274

Search completed: August 2, 2004, 14:58:33
Job time : 13 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:54:48 ; Search time 36.8 Seconds
(without alignments)
1287.123 Million cell updates/sec

Title: US-09-543-407-18
Perfect score: 780
Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMRVQVGFNNATNQY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	520	66.7	151	12	US-09-741-873B-4
2	520	66.7	151	12	US-09-741-873B-4
3	444	56.9	131	12	US-09-741-873B-2
4	444	56.9	131	12	US-09-741-873B-2
5	108.5	13.9	445	15	US-10-369-493-20638
6	93	11.9	974	12	US-10-282-122A-44999
7	91.5	11.7	580	12	US-10-647-057-4
8	91	11.7	597	9	US-09-793-306-146
9	90	11.5	906	12	US-10-233-449-7
10	90	11.5	906	14	US-10-251-661-2
11	89	11.4	1448	16	US-10-408-765A-998
12	88.5	11.3	204	12	US-10-424-599-203972
13	88.5	11.3	244	12	US-10-425-114-37715
14	88.5	11.3	253	16	US-10-437-963-114193
15	88.5	11.3	283	12	US-10-424-599-232271

16	88	11.3	3619	12	US-10-282-122A-67768
17	88	11.3	6310	12	US-10-282-122A-67793
18	87.5	11.2	172	12	US-10-425-114-46013
19	86.5	11.1	208	16	US-10-475-970-16
20	85.5	11.0	278	9	US-09-810-264-28
21	85.5	11.0	588	12	US-10-282-122A-64869
22	84.5	10.8	408	15	US-10-369-493-12833
23	84.5	10.8	847	16	US-10-437-963-118741
24	84	10.8	283	12	US-10-424-599-168890
25	84	10.8	892	10	US-09-952-267-5
26	84	10.8	943	9	US-09-996-634-131
27	84	10.8	943	10	US-09-997-182-131
28	84	10.8	943	10	US-09-997-181-131
29	84	10.8	943	14	US-10-193-002-199
30	84	10.8	943	14	US-10-084-843-204
31	84	10.8	1246	12	US-10-282-122A-49773
32	84	10.8	3300	12	US-10-282-122A-64369
33	83.5	10.7	65	9	US-09-996-194-16
34	83.5	10.7	65	12	US-10-164-966-33
35	83.5	10.7	628	12	US-10-282-122A-53269
36	83.5	10.7	873	10	US-09-952-267-13
37	83	10.6	182	14	US-10-238-075-1549
38	83	10.6	354	10	US-09-828-843A-21
39	83	10.6	518	9	US-09-976-297-2
40	83	10.6	720	9	US-09-801-368-176
41	83	10.6	961	12	US-10-282-122A-76322
42	83	10.6	1306	12	US-10-282-122A-64405
43	83	10.6	2353	9	US-09-797-862-33
44	82.5	10.6	434	15	US-10-045-674-594
45	82.5	10.6	574	12	US-10-282-122A-55429

ALIGNMENTS

RESULT 1

US-09-741-873B-4
; Sequence 4; Application US/09741873B
; Publication No. US20020081722A1
; GENERAL INFORMATION:
; APPLICANT: Olsear, Arne
; APPLICANT: Normark, Staffan
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; CURRENT FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1994-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-4

Query Match 66.7%; Score 520; DB 12; Length 151;
Best Local Similarity 68.2%; Pred. No. 2.2e+45;
Matches 103; Conservative 21; Mismatches 27; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVVSGSALGVVPPQWGGGNNHGGSSGPDYDQIVTVVTHMAHALQ 60

```

Db      1  MKLLKVAATAAIVFSGSAGVVPQYGGGNGHGGGNSGPNSELNIYQYGGNSALALQ 60
      61  SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNNAITDQWNAKNSDITVQYGG 120
      61  TDARNSDLITQHGNGADVGQSDSSIDITQRFNGSATLDQWNGKNSMTVKQFG 120
      121  NNAALVNQASDSSVMVRQVGFNNATANQY 151
      121  GNGAAVDQASNSVNVTVQFGNNATAHQY 151

```

RESULT 2

```

US-09-741-873B-4
; Sequence 4, Application US/09741873B
; Publication No. US20040096965A9
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; PRIOR FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli

```

```

Query Match          66.7%; Score 520; DB 12; Length 151;
Best Local Similarity 68.2%; Pred. No. 2.2e-45;
Matches 103; Conservative 21; Mismatches 27; Indels 0; Gaps 0;

QY      1  MKLLKVAATAAIVFSGSAGVVPQYGGGNGHGGGNSGPNSELNIYQYGGNSALALQ 60
      61  SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNNAITDQWNAKNSDITVQYGG 120
      61  TDARNSDLITQHGNGADVGQSDSSIDITQRFNGSATLDQWNGKNSMTVKQFG 120
      121  NNAALVNQASDSSVMVRQVGFNNATANQY 151
      121  GNGAAVDQASNSVNVTVQFGNNATAHQY 151

```

RESULT 3

```

US-09-741-873B-2
; Sequence 2, Application US/09741873B
; Publication No. US20020081722A1
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; CURRENT FILING DATE: 2003-04-04

```

```

Query Match          66.7%; Score 520; DB 12; Length 151;
Best Local Similarity 68.2%; Pred. No. 2.2e-45;
Matches 103; Conservative 21; Mismatches 27; Indels 0; Gaps 0;

QY      1  MKLLKVAATAAIVFSGSAGVVPQYGGGNGHGGGNSGPNSELNIYQYGGNSALALQ 60
      61  SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNNAITDQWNAKNSDITVQYGG 120
      61  TDARNSDLITQHGNGADVGQSDSSIDITQRFNGSATLDQWNGKNSMTVKQFG 120
      121  NNAALVNQASDSSVMVRQVGFNNATANQY 151
      121  GNGAAVDQASNSVNVTVQFGNNATAHQY 151

```

RESULT 4

```

US-09-741-873B-2
; Sequence 2, Application US/09741873B
; Publication No. US20040096965A9
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; CURRENT FILING DATE: 2003-04-04

```

```

Query Match          56.9%; Score 444; DB 12; Length 131;
Best Local Similarity 65.6%; Pred. No. 1.12e-37;
Matches 86; Conservative 19; Mismatches 26; Indels 0; Gaps 0;

QY      21  GVVPQWGGGNGHGGGNSGPNSELNIYQYGGNSALALQTDARNSDLITQHGNGAD 80
      61  GVVPQWGGGNGHGGGNSGPNSELNIYQYGGNSALALQTDARNSDLITQHGNGAD 60
      81  VQGDADNSTIETQNGFRNNAITDQWNAKNSDITVQYGGNNAALVNQASDSSVMVRQV 140
      61  VQGDSDSSIDITQRFNGSATLDQWNGKNSMTVKQFGGNGAAVDQASNSVNVTVQV 120
      141  GFGNNATANQY 151
      121  GFGNNATAHQY 131

```

```

; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-2

```

```

Query Match          56.9%; Score 444; DB 12; Length 131;
Best Local Similarity 65.6%; Pred. No. 1.12e-37;
Matches 86; Conservative 19; Mismatches 26; Indels 0; Gaps 0;

QY      21  GVVPQWGGGNGHGGGNSGPNSELNIYQYGGNSALALQTDARNSDLITQHGNGAD 80
      61  GVVPQWGGGNGHGGGNSGPNSELNIYQYGGNSALALQTDARNSDLITQHGNGAD 60
      81  VQGDADNSTIETQNGFRNNAITDQWNAKNSDITVQYGGNNAALVNQASDSSVMVRQV 140
      61  VQGDSDSSIDITQRFNGSATLDQWNGKNSMTVKQFGGNGAAVDQASNSVNVTVQV 120
      141  GFGNNATANQY 151
      121  GFGNNATAHQY 131

```

RESULT 4

```

US-09-741-873B-2
; Sequence 2, Application US/09741873B
; Publication No. US20040096965A9
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; CURRENT FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-2

```

Query Match

56.9%; Score 444; DB 12; Length 131;

```
Best Local Similarity 65.6%; Pred. No. 1.2e-37;
Matches 86; Conservative 19; Mismatches 26; Indels 0; Gaps 0;

QY 21 GWVPWGGGNNHGGSSGPDYDQLVTVRVVTHEMAHALQSDARKSETTITQSGYGNAD 80
Db 1 GWVPYGGGNGGGGNSGSELNTVYGGGNSALALQTDARNSLDITQHGCGNGAD 60
QY 81 VGGADNSTIELTQGFNNATIDQWNAKSDITVGYGGNNAALVNQTSDDSSVMYQV 140
Db 61 VQGSDDSSIDLTRGFNSATLDQWNGKSEMTVKQFGGNGAAYDQTSNNSVNTQV 120
QY 141 GFGNNATANQY 151
Db 121 GFGNNATAHQY 131

RESULT 5
US-10-369-493-20638
; Sequence 20638, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 20638
; LENGTH: 445
; TYPE: PRT
; ORGANISM: Rhodospseudomonas palustris
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(445)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-20638

Query Match 13.9%; Score 108.5; DB 15; Length 445;
Best Local Similarity 27.5%; Pred. No. 0.015;
Matches 46; Conservative 19; Mismatches 61; Indels 41; Gaps 6;

QY 7 AAFRA-----IVVSGSALAGVVPWQGGG-----NHNG-----GNSGPDYDQLVTRV 50
Db 19 AAFADSNTVYLNQDGNQANITQSGNGNSVGAFNGSGFLQENGTLSGAN---LLT-- 73
QY 51 VTHEMAHALQSDARKSETTITQSGYGNADYVGGADNSTIELTQGFNNATIDQWNAKN 110
Db 74 ---VKQSGNSVGRDITQKQSGAGNSAAIFQETGSDVDELQQTGTSNGAVPSGWNWN 129
QY 111 -----SDITVGYGGNNAALVNQTSDDSSVMYQVQ 141
Db 130 DPGVFNKITQDSSNSGSKSVIQDGKNVFSIKQNTGNSTSVNQIG 176

RESULT 6
US-10-282-122A-44999
; Sequence 44999, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
```

```
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 44999
; LENGTH: 974
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-10-282-122A-44999

Query Match 11.9%; Score 93; DB 12; Length 974;
Best Local Similarity 24.5%; Pred. No. 1.5;
Matches 36; Conservative 23; Mismatches 52; Indels 36; Gaps 7;

QY 15 SGSALAGVVPWQGGGNNHGG-GNSSGPDYDQLVTVRVVTHEM-----AHALQSDA 63
Db 299 AGNGIA-----SGNGEYHVGNGNGDDVD--ITAPITGVNLISGNSFTLGNSSSSV 350
QY 64 RKSETTITQS-----GYGNGADVGGADNSTIELTQNGF-----RNNATIDQWNAKNS 111
Db 351 NTAPITTSNTVNDNDIDNGNSGTTGSGSGNGSGDGLLNGAASGNGEYHVGNGNGDDV 410
QY 112 DIT-----VGQYGGNNAALVNQTSADS 133
Db 411 DITAPITGVNFGNSFSLIGNSSSS 437

RESULT 7
US-10-647-057-4
; Sequence 4; Application US/10647057
; Publication No. US20040047871A1
; GENERAL INFORMATION:
; APPLICANT: KANSAS STATE UNIVERSITY RESEARCH FOUNDATION
; APPLICANT: NAGARAJA, T.
; APPLICANT: STEWART, GEORGE
; APPLICANT: NARAYANAN, SANJEEV
; APPLICANT: CHENGAPPA, M.
; TITLE OF INVENTION: RECOMBINANT FUSOBACTERIUM NECROPHORUM LEUKOTOXIN VACCINE AND PRE
; FILE REFERENCE: PCT-30962
; CURRENT APPLICATION NUMBER: US/10/647,057
; CURRENT FILING DATE: 2003-08-22
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 4
```


QY 56 AHALQSDARKSETTITQSGYNGADV-----GQGAD-----NSTIELTQ 94
Db 301 AAFOSLRQ---RIDISRRGNAGDCLANPAVPMGQIDIGRALQQVAFELTGNVQFNE 357
QY 95 NGFRNNAT-----IDOWNAKNSDI---TVQYGGNNAALVNQT-----ASP 132
Db 358 KGRRTNYTLHVIEKHGIRKIGYWNEDKFXPAATDAQAGDNSSVQNRYYIVTTILED 417
QY 133 SSVVMVRQVFGNNATANQY 151
Db 418 PYWMLK-----NANQF 429

RESULT 11

US-10-408-765A-998
; Sequence 998, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Choeh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 998
; LENGTH: 1448
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-998

Query Match 11.4%; Score 89; DB 16; Length 1448;
Best Local Similarity 24.4%; Pred. No. 6.5;
Matches 39; Conservative 18; Mismatches 47; Indels 56; Gaps 7;
QY 10 AAIIVVSGSALAGV-----PWG-----GGN-----H 32
Db 307 AAAAKSGHAWGAAANQEDKSPWGPBPKPSQHWGQORSNPAWSAGGDWADSSVLGH 366
QY 33 NCGGNSGPDYDQLVTRVVTTHMAHALQSDARKSETTIT-OSGYNGADVCGGADNSTIE 91
Db 367 LGDGKKGSGMD-----ADSNRSGSGWNDTTRSGNSGWSNSTNTKANPCTNWGE 415
QY 92 LTQNGFRNNATIDOWNAKNSDITVQYGGNNAALVNQTAS 131
Db 416 TLKPGPQN-----NASKFQDNVSVNWGG--AASVKQTGT 448

RESULT 12

US-10-424-599-203972
; Sequence 203972, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 203972
; LENGTH: 204
; TYPE: PRT

; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_26213C.1.pcp
US-10-424-599-203972
Query Match 11.3%; Score 88.5; DB 12; Length 204;
Best Local Similarity 25.7%; Pred. No. 0.61;
Matches 35; Conservative 10; Mismatches 30; Indels 61; Gaps 7;
QY 21 GVVFPQWGGGNGHNGGNSGPDYDQLVTRVVTTHMAHALQSDARKSETTITQSGYNGAD 80
Db 106 GRVP-W-GGNSHERGGYSDG-----NSD-----SGWQGGQD 134
QY 81 VGOGADNSTIELTQNGFRNNATIDOWNAKNSDIT-----VGOYGGNNAALVNQTAS 131
Db 135 QGSGSDG-----NGQGRGWRGNGNNSERNKDEESRGNWSGFGNAGSGNENS- 182
QY 132 DSSVMVRQVFGNNAT 147
Db 183 -----GWGKNAT 189

RESULT 13

US-10-425-114-37715
; Sequence 37715, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 37715
; LENGTH: 244
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3170-053-G12_F11.pcp
US-10-425-114-37715

Query Match 11.3%; Score 88.5; DB 12; Length 244;
Best Local Similarity 22.3%; Pred. No. 0.76;
Matches 32; Conservative 21; Mismatches 48; Indels 39; Gaps 6;
QY 27 GGGGNGGNGGNSGPDY-----DQLVTRVVTTHMAHALQSDARKSETTITQSG- 74
Db 23 GGGGNNNNNNNGTNYFATSSAPSLFSNSVNASALSHMSATALLQKAAQMGATTNGGT 82
QY 75 -----YNGADVCGGADNSTIELTQNGFRNNATIDOWNAKNSDITVQYGGNNAALVN- 127
Db 83 ASLLKSPGSSSSSGGSKLV-----NAAN--YVSGMFGNNH--VNE 122
QY 128 QTASDSSVMVRQVFGNNAT 147
Db 123 QSNLNQDLNNSFAVGNGSS 142

RESULT 14

US-10-437-963-114193
; Sequence 114193, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei

Search completed: August 2, 2004, 15:36:11
Job time : 37.8 secs

; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 114193
; LENGTH: 253
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_17908C.1.pdp
US-10-437-963-114193

Query Match 11.3%; Score 88.5; DB 16; Length 253;
Best Local Similarity 26.6%; Pred. No. 0.8;
Matches 38; Conservative 14; Mismatches 48; Indels 43; Gaps 6;

QY 4 LKVA--PRAIVVSSGALAGVVPWG-----GGGNHGGG-----NSSGP 41
DB 65 IKIALGFVLLSIGLASARVERYSSESGTNGEGGVYNGGVSGSGAGSGSGG 124
QY 42 DYDQLVTRVVTHEMAHALQSDARKSETTITQSGY---GNGADVCGGADNSTIETLQNGF 97
DB 125 NYG-----AHASGGGGEG-----GGYQYGGSGSGSGSGSSEYTONGG 167
QY 98 RNNATIDQNAKNSDITVQYGG 120
DB 168 YGYAGSSAGSGGAGAGAGAGG 190

RESULT 15
US-10-424-599-232271
; Sequence 232271, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 232271
; LENGTH: 283
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_51764C.1.pdp
US-10-424-599-232271

Query Match 11.3%; Score 88.5; DB 12; Length 283;
Best Local Similarity 22.9%; Pred. No. 0.92;
Matches 32; Conservative 21; Mismatches 48; Indels 39; Gaps 6;

QY 27 GGGNHGGGSGSPDY-----DQLVTRVVTHEMAHALQSDARKSETTITQSG- 74
DB 62 GGGNNNNNNNEGTYFATSPSLFSNVNASLSHMSATALLQKAAQMGATTSNGGT 121
QY 75 -----YNGADVGGGADNSTIETLQNGFRNATIDQNAKNSDITVQYCGNNAALVN- 127
DB 122 ASLLKSFSGASSSGSGSGSKLV-----NAAN--YVGMFGGNH--VNE 161
QY 128 QTASDSVMVQVGFQNNAT 147
DB 162 QSNSNLQDLNLSFVGGNKS 181

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:48:33 ; Search time 167.9 Seconds
(without alignments)
877.809 Million cell updates/sec

Title: US-09-543-407-18
Perfect score: 780
Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVFGNNATANQY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 6019581 seqs, 97603577 residues

Total number of hits satisfying chosen parameters: 6019581

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending Parents AA Main:*

1: /cgn2_6/ptodata/2/paa/PCTUS_COMB.pep.*
2: /cgn2_6/ptodata/2/paa/US06_COMB.pep.*
3: /cgn2_6/ptodata/2/paa/US07_COMB.pep.*
4: /cgn2_6/ptodata/2/paa/US080_COMB.pep.*
5: /cgn2_6/ptodata/2/paa/US081_COMB.pep.*
6: /cgn2_6/ptodata/2/paa/US082_COMB.pep.*
7: /cgn2_6/ptodata/2/paa/US083_COMB.pep.*
8: /cgn2_6/ptodata/2/paa/US084_COMB.pep.*
9: /cgn2_6/ptodata/2/paa/US085_COMB.pep.*
10: /cgn2_6/ptodata/2/paa/US086_COMB.pep.*
11: /cgn2_6/ptodata/2/paa/US087_COMB.pep.*
12: /cgn2_6/ptodata/2/paa/US088_COMB.pep.*
13: /cgn2_6/ptodata/2/paa/US089_COMB.pep.*
14: /cgn2_6/ptodata/2/paa/US090_COMB.pep.*
15: /cgn2_6/ptodata/2/paa/US091_COMB.pep.*
16: /cgn2_6/ptodata/2/paa/US092_COMB.pep.*
17: /cgn2_6/ptodata/2/paa/US093_COMB.pep.*
18: /cgn2_6/ptodata/2/paa/US094_COMB.pep.*
19: /cgn2_6/ptodata/2/paa/US095_COMB.pep.*
20: /cgn2_6/ptodata/2/paa/US096_COMB.pep.*
21: /cgn2_6/ptodata/2/paa/US097A_COMB.pep.*
22: /cgn2_6/ptodata/2/paa/US097B_COMB.pep.*
23: /cgn2_6/ptodata/2/paa/US098_COMB.pep.*
24: /cgn2_6/ptodata/2/paa/US099A_COMB.pep.*
25: /cgn2_6/ptodata/2/paa/US099B_COMB.pep.*
26: /cgn2_6/ptodata/2/paa/US100_COMB.pep.*
27: /cgn2_6/ptodata/2/paa/US101_COMB.pep.*
28: /cgn2_6/ptodata/2/paa/US102_COMB.pep.*
29: /cgn2_6/ptodata/2/paa/US103_COMB.pep.*
30: /cgn2_6/ptodata/2/paa/US104_COMB.pep.*
31: /cgn2_6/ptodata/2/paa/US105_COMB.pep.*
32: /cgn2_6/ptodata/2/paa/US107_COMB.pep.*
33: /cgn2_6/ptodata/2/paa/US60_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
------------	-------	-------------	--------	-------	-------------

1	780	100.0	151	19	US-09-543-407-18	Sequence 18, Appl
2	691	87.6	151	19	US-09-543-407-5	Sequence 5, Appl
3	686	87.9	151	6	US-08-233-642A-57	Sequence 57, Appl
4	675	86.5	151	19	US-09-543-407-36	Sequence 26, Appl
5	622	79.7	151	19	US-09-543-407-20	Sequence 20, Appl
6	618	79.2	151	19	US-09-543-407-16	Sequence 16, Appl
7	613	78.6	151	19	US-09-543-407-12	Sequence 12, Appl
8	611	78.3	151	19	US-09-543-407-14	Sequence 14, Appl
9	608	77.9	151	19	US-09-543-407-24	Sequence 24, Appl
10	604	77.4	131	19	US-09-543-407-31	Sequence 31, Appl
11	601	77.1	151	19	US-09-543-407-28	Sequence 28, Appl
12	600	76.9	151	19	US-09-543-407-22	Sequence 22, Appl
13	599	76.8	151	19	US-09-543-407-30	Sequence 30, Appl
14	523	67.1	151	19	US-09-543-407-7	Sequence 7, Appl
15	520	66.7	151	13	US-08-978-878-4	Sequence 4, Appl
16	520	66.7	151	21	US-09-741-873B-4	Sequence 4, Appl
17	518	66.4	151	33	US-60-352-946-2	Sequence 2, Appl
18	518	66.4	151	33	US-60-444-371-2	Sequence 2, Appl
19	506	64.9	120	6	US-08-233-642A-55	Sequence 55, Appl
20	483	61.3	109	19	US-09-543-407-34	Sequence 34, Appl
21	457	58.6	158	16	US-09-252-691C-5834	Sequence 5834, Ap
22	457	58.6	158	16	US-09-252-691C-5834	Sequence 5834, Ap
23	457	58.6	158	30	US-10-417-886-5834	Sequence 5834, Ap
24	444	56.9	131	13	US-08-978-878-2	Sequence 2, Appl
25	444	56.9	131	21	US-09-741-873B-2	Sequence 2, Appl
26	359.5	46.1	109	19	US-09-543-407-35	Sequence 35, Appl
27	276	35.4	68	19	US-09-543-407-37	Sequence 37, Appl
28	250.5	32.1	70	19	US-09-543-407-32	Sequence 32, Appl
29	175	22.4	48	19	US-09-543-407-39	Sequence 39, Appl
30	116	14.9	186	16	US-09-252-691C-5833	Sequence 5833, Ap
31	116	14.9	186	16	US-09-252-691C-5833	Sequence 5833, Ap
32	116	14.9	186	30	US-10-417-886-5833	Sequence 5833, Ap
33	108.5	13.9	445	29	US-10-369-493-20638	Sequence 20638, A
34	108.5	13.9	445	29	US-10-369-493-20638	Sequence 20638, A
35	107.5	13.8	151	19	US-09-543-407-6	Sequence 6, Appl
36	106.5	13.7	151	19	US-09-543-407-6	Sequence 6, Appl
37	101	12.9	252	33	US-60-173-464-21553	Sequence 21553, A
38	101	12.9	252	30	US-09-614-150-25818	Sequence 25818, A
39	101	12.9	252	20	US-09-614-150A-25818	Sequence 25818, A
40	101	12.9	252	33	US-60-191-637-25957	Sequence 25957, A
41	101	12.9	252	33	US-60-191-637-25957	Sequence 25957, A
42	100	12.8	1308	27	US-10-179-131-5148	Sequence 5148, Ap
43	99.5	12.8	1249	30	US-10-455-719-358	Sequence 358, App
44	99.5	12.8	1249	33	US-60-385-568-357	Sequence 357, App
45	99.5	12.8	1249	33	US-60-446-775-358	Sequence 358, App

ALIGNMENTS

RESULT 1
US-09-543-407-18
; Sequence 18, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIBRILLAR SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.

US-09-543-407-18

Query Match 100.0%; Score 780; DB 19; Length 151;
 Best Local Similarity 100.0%; Pred. No. 3.7e-75;
 Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNNHNGGNSGPDYDQLVTRVVTHEMAHALQ 60
 |||||
 Db 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNNHNGGNSGPDYDQLVTRVVTHEMAHALQ 60
 |||||

Qy 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
 |||||
 Db 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
 |||||

Qy 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
 |||||
 Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
 |||||

RESULT 2

US-09-543-407-5
 ; Sequence 5, Application US/09543407
 ; GENERAL INFORMATION:
 ; APPLICANT: White, Aaron P.
 ; APPLICANT: Doran, James L.
 ; APPLICANT: Collinson, S. Karen
 ; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
 ; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
 ; FILE REFERENCE: 920043.406
 ; CURRENT APPLICATION NUMBER: US/09/543,407
 ; CURRENT FILING DATE: 2000-04-05
 ; NUMBER OF SEQ ID NOS: 59
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 5
 ; LENGTH: 151
 ; TYPE: PRT
 ; ORGANISM: Salmonella enteritidis
 US-09-543-407-5

Query Match 88.6%; Score 691; DB 19; Length 151;
 Best Local Similarity 90.7%; Pred. No. 1.4e-65;
 Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

Qy 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNNHNGGNSGPDYDQLVTRVVTHEMAHALQ 60
 |||||
 Db 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQYGSANAALALQ 60
 |||||

Qy 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
 |||||
 Db 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
 |||||

Qy 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
 |||||
 Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
 |||||

RESULT 3

US-08-233-642A-57
 ; Sequence 57, Application US/08233642A
 ; GENERAL INFORMATION:
 ; APPLICANT: Kay, William W.
 ; APPLICANT: Collinson, S. Karen
 ; APPLICANT: Clouthier, Sharon C.
 ; APPLICANT: Doran, James L.
 ; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
 ; BASED VACCINES
 ; NUMBER OF SEQUENCES: 58
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Seed and Berry
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue
 ; CITY: Seattle

STATE: Washington
 COUNTRY: U.S.A.
 ZIP: 98104-7092
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/233,642A
 FILING DATE: 26-APR-1994
 CLASSIFICATION: 424
 ATTORNEY/AGENT INFORMATION:
 NAME: King, Joshua
 REGISTRATION NUMBER: 35,570
 REFERENCE/DOCKET NUMBER: 920043.403C3
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 622-4900
 TELEFAX: (206) 682-6031
 TELEX: 3723836 SEEDANBERRY
 INFORMATION FOR SEQ ID NO: 57:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 151 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-233-642A-57

Query Match 87.9%; Score 686; DB 6; Length 151;
 Best Local Similarity 90.1%; Pred. No. 4.8e-65;
 Matches 136; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

Qy 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNNHNGGNSGPDYDQLVTRVVTHEMAHALQ 60
 |||||
 Db 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQYGSANAALALQ 60
 |||||

Qy 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
 |||||
 Db 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
 |||||

Qy 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
 |||||
 Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
 |||||

RESULT 4
 US-09-543-407-26
 ; Sequence 26, Application US/09543407
 ; GENERAL INFORMATION:
 ; APPLICANT: White, Aaron P.
 ; APPLICANT: Doran, James L.
 ; APPLICANT: Collinson, S. Karen
 ; APPLICANT: Kay, William W.
 ; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
 ; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
 ; FILE REFERENCE: 920043.406
 ; CURRENT APPLICATION NUMBER: US/09/543,407
 ; CURRENT FILING DATE: 2000-04-05
 ; NUMBER OF SEQ ID NOS: 59
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 26
 ; LENGTH: 151
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Recombinant Salmonella enteritidis 2b afga
 ; OTHER INFORMATION: sequence containing the replacement fragment
 ; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
 US-09-543-407-26

Query Match 86.5%; Score 675; DB 19; Length 151;
 Best Local Similarity 81.9%; Pred. No. 7.3e-64;
 Matches 136; Conservative 0; Mismatches 0; Indels 30; Gaps 2;

QY 1 M K L L K V A A F A A I V V S G S A L A G V V P Q W G G G N H G G N S S G P D ----- Y D Q 45
D b 1 M K L L K V A A F A A I V V S G S A L A G V V P Q W G G G N H G G N S S G P D S T L S I Y Q Y G S A N A A L Q 60
Q Y 46 L V T R V T V T H E M A H A L Q S D A R K S E T T I T Q S G Y G N G A D V G Q G A D N S T I E L T Q N G F R N N A T I D Q 105
D b 61 L V T R V T V T H E M A H A ----- G Y G N G A D V G Q G A D N S T I E L T Q N G F R N N A T I D Q 105
Q Y 106 W N A K N S D I T V G Q Y G G N N A A L V N Q T A S D S S V M V R Q V G F G N N A T A N Q Y 151
D b 106 W N A K N S D I T V G Q Y G G N N A A L V N Q T A S D S S V M V R Q V G F G N N A T A N Q Y 151

RESULT 5
US-09-543-407-20
; Sequence 20, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-20

Query Match 79.7%; Score 622; DB 19; Length 151;
Best Local Similarity 73.6%; Pred. No. 3.6e-58;
Matches 128; Conservative 0; Mismatches 0; Indels 46; Gaps 2;
QY 1 M K L L K V A A F A A I V V S G S A L A G V V P Q W G G G N H G G N S S G P D ----- 42
D b 1 M K L L K V A A F A A I V V S G S A L A G V V P Q W G G G N H G G N S S G P D S T L S I Y Q Y G S A N A A L Q 60
Q Y 43 ----- Y D Q L V T R V T H E M A H A L Q S D A R K S E T T I T Q S G Y G N G A D V G Q G A D N S T I E L T Q N G F 97
D b 61 S D A R K Y D Q L V T R V T H E M A H A ----- G Q G A D N S T I E L T Q N G F 97
Q Y 98 R N N A T I D O W N A K N S D I T V G Q Y G G N N A A L V N Q T A S D S S V M V R Q V G F G N N A T A N Q Y 151
D b 98 R N N A T I D O W N A K N S D I T V G Q Y G G N N A A L V N Q T A S D S S V M V R Q V G F G N N A T A N Q Y 151

RESULT 6
US-09-543-407-16
; Sequence 16, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 151

; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-16

Query Match 79.2%; Score 618; DB 19; Length 151;
Best Local Similarity 76.6%; Pred. No. 9.8e-58;
Matches 131; Conservative 0; Mismatches 0; Indels 40; Gaps 2;
QY 1 M K L L K V A A F A A I V V S G S A L A G V V P Q W G G G N H G G N S S G P D Y D Q L V T R V T H E M A H A --- 57
D b 1 M K L L K V A A F A A I V V S G S A L A G V ----- Y D Q L V T R V T H E M A H A S G 40
QY 58 ----- A L Q S D A R K S E T T I T Q S G Y G N G A D V G Q G A D N S T I E L T Q N G F R N N 100
D b 41 P D S T L S I Y Q Y G S A N A A L Q S D A R K S E T T I T Q S G Y G N G A D V G Q G A D N S T I E L T Q N G F R N N 100
QY 101 A T I D O W N A K N S D I T V G Q Y G G N N A A L V N Q T A S D S S V M V R Q V G F G N N A T A N Q Y 151
D b 101 A T I D O W N A K N S D I T V G Q Y G G N N A A L V N Q T A S D S S V M V R Q V G F G N N A T A N Q Y 151

RESULT 7
US-09-543-407-12
; Sequence 12, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-12

Query Match 78.6%; Score 613; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 3.4e-57;
Matches 122; Conservative 7; Mismatches 22; Indels 0; Gaps 0;
QY 1 M K L L K V A A F A A I V V S G S A L A G V V P Q W G G G N H G G N S S G P D Y D Q L V T R V T H E M A H A L Q 60
D b 1 M K L L K V A A F A A I V V S G S A L A G V V P Q W G G G N H G G N S S G P D S T L S I Y Q Y G S A N A A L Q 60
QY 61 S D A R K S E T T I T Q S G Y G N G A D V G Q G A D N S T I E L T Q N G F R N N A T I D O W N A K N S D I T V G Q Y G G 120
D b 61 S D A R K S E T T I T Q S G Y G N G A D V G Q G A D N S T I E L T Q N G F R N N A T I D O W N A K N S D I T V G Q Y G G 120
QY 121 N N A A L V N Q T A S D S S V M V R Q V G F G N N A T A N Q Y 151
D b 121 N N A A L V N Q L V T R V T H E M A H A N N A T A N Q Y 151

RESULT 8
US-09-543-407-14
; Sequence 14, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.

```
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-14

Query Match      78.3%; Score 611; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 5.5e-57;
Matches 123; Conservative 6; Mismatches 22; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPWGGGNGHNGSGGPDYDQVTRVVTHEMAHALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPWGGGNGHNGSGGPDSTLSIYQYGSANAALQ 60

QY 61 SPARKSETTITQSGYNGADVQGGADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGG 120
DB 61 SPARKSETTITQSGYNGADVQGGADNSTIELTQNGFRNNATIDQWNAKNSDITVQYDQ 120

QY 121 NNAALVNOTASDSSVMVROVQVFGNNATANQY 151
DB 121 LVTRVVTHEMAHASVMVROVQVFGNNATANQY 151

RESULT 9
US-09-543-407-24
; Sequence 24, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-24

Query Match      77.9%; Score 608; DB 19; Length 151;
Best Local Similarity 82.1%; Pred. No. 1.2e-56;
Matches 124; Conservative 5; Mismatches 22; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPWGGGNGHNGSGGPDYDQVTRVVTHEMAHALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPWGGGNGHNGSGGPDSTLSIYQYGSANAALQ 60

QY 61 SPARKSETTITQSGYNGADVQGGADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGG 120
DB 61 SPARKSETTITQSGYNGADVQGGADNSTIELTQNGFRNNATIDQWNAKNSDITVTRVVT 120
```

```
QY 121 NNAALVNOTASDSSVMVROVQVFGNNATANQY 151
DB 121 HEMAHANQTASDSSVMVROVQVFGNNATANQY 151

RESULT 10
US-09-543-407-31
; Sequence 31, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Salmonella enteritidis
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-31

Query Match      77.4%; Score 604; DB 19; Length 131;
Best Local Similarity 89.3%; Pred. No. 2.6e-56;
Matches 117; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 21 GVVPOWGGGNGHNGSGGPDYDQVTRVVTHEMAHALQSDARKSETTITQSGYNGAD 80
DB 1 GVVPOWGGGNGHNGSGGPDSTLSIYQYGSANAALQSDARKSETTITQSGYNGAD 60

QY 81 VQGGADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGSANAALVNOTASDSSVMVROV 140
DB 61 VQGGADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGSANAALVNOTASDSSVMVROV 120

QY 141 GFGNNATANQY 151
DB 121 GFGNNATANQY 131

RESULT 11
US-09-543-407-28
; Sequence 28, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-28

Query Match      77.1%; Score 601; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 6.6e-56;
Matches 122; Conservative 4; Mismatches 25; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPWGGGNGHNGSGGPDYDQVTRVVTHEMAHALQ 60
```

```
Db 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADYVQGGADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120
Db 61 SDARKSETTITQSGYNGADYVQGGADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120
QY 121 NNAALVNQTSASSVMVROVGFNNATANQY 151
Db 121 NNAALVNQTSASSVMVROVGFNNATANQY 151
```

RESULT 12

```
US-09-543-407-22
; Sequence 22, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 22
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-22
```

```
Query Match 76.9%; Score 600; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 8.4e-56;
Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQGSANAALALQ 60
Db 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADYVQGGADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120
Db 61 SDARKSETTITQSGYNGADYVQGGADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120
QY 121 NNAALVNQTSASSVMVROVGFNNATANQY 151
Db 121 NNAALVNQTSASSVMVROVGFNNATANQY 151
```

RESULT 13

```
US-09-543-407-30
; Sequence 30, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
```

```
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-30

Query Match 76.8%; Score 599; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 1.1e-55;
Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQGSANAALALQ 60
Db 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADYVQGGADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120
Db 61 SDARKSETTITQSGYNGADYVQGGADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120
QY 121 NNAALVNQTSASSVMVROVGFNNATANQY 151
Db 121 NNAALVNQTSASSVMVROVGFNNATANQY 151
```

RESULT 14

```
US-09-543-407-7
; Sequence 7, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-543-407-7
```

```
Query Match 67.1%; Score 523; DB 19; Length 151;
Best Local Similarity 68.9%; Pred. No. 1.6e-47;
Matches 104; Conservative 20; Mismatches 27; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQGSANAALALQ 60
Db 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADYVQGGADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120
Db 61 TDARNSDLTITQHGNGGADYVQGGDDSDSIDTQRGFGNSATLDQNGKNSMTVKQFGG 120
QY 121 NNAALVNQTSASSVMVROVGFNNATANQY 151
Db 121 NGCAAVDQTSASSVMVROVGFNNATANQY 151
```

RESULT 15

```
US-08-978-878-4
; Sequence 4, Application US/08978878
; GENERAL INFORMATION:
; APPLICANT: NORMARK, Staffan
; APPLICANT: OLSEN, Arne
; TITLE OF INVENTION: FIBRONECTIN BINDING PROTEIN AS WELL AS ITS PREPARATION
; FILE REFERENCE: 012889-081
; CURRENT APPLICATION NUMBER: US/08/978,878
; CURRENT FILING DATE: 1997-11-26
; EARLIER APPLICATION NUMBER: SE 8801723-1
; EARLIER FILING DATE: 1988-05-06
```

```
; EARLIER APPLICATION NUMBER: US 07/347,189
; EARLIER FILING DATE: 1989-05-04
; EARLIER APPLICATION NUMBER: US 07/789,437
; EARLIER FILING DATE: 1991-11-06
; EARLIER APPLICATION NUMBER: US 07/970,846
; EARLIER FILING DATE: 1992-11-03
; EARLIER APPLICATION NUMBER: US 08/187,865
; EARLIER FILING DATE: 1994-01-28
; EARLIER APPLICATION NUMBER: US 08/318,519
; EARLIER FILING DATE: 1994-10-05
; EARLIER APPLICATION NUMBER: US 08/495,959
; EARLIER FILING DATE: 1995-06-28
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-08-978-878-4

Query Match          66.7%; Score 520; DB 13; Length 151;
Best Local Similarity 68.2%; Pred. No. 3.4e-47;
Matches 103; Conservative 21; Mismatches 27; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNNHGGSSGPPDYDOLVTRVVTHEMAHALQ 60
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1 MKLLKVAATAAIVFSGSAVAGVVPQYGGGNNHGGSSGPPSENLTYQYGGNSALALQ 60
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 61 SDARKSETTITOSGYGNGADVCGGADNSTIELTQNGFRNATIDOWNAKNSDITVGOYGG 120
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 61 TDARNSDLTITQGGGNGADVCGGDDSSIDLTRQFGNSATLDWNGKNSMTVKQFEGG 120
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 121 NNAALVNOTASDSVMVRQVGFNNATANQY 151
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 121 GNGAAVDQTASNSVNVTVQVGFNNATAHQY 151
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
```

Search completed: August 2, 2004, 15:26:43
Job time : 167.9 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:49:38 ; Search time 17.8 Seconds
(without alignments)
888.146 Million cell updates/sec

Title: US-09-543-407-18
Perfect score: 780
Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSWVTVQVGFNNATANYQ 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 601315 seqs, 104695340 residues

Total number of hits satisfying chosen parameters: 601315

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_AA_New:*
1: /cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pep:*
2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep:*
3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep:*
4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep:*
5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:*
6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:*
7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	520	66.7	151	US-09-741-873C-4	Sequence 4, Appli
2	444	56.9	131	US-09-741-873C-2	Sequence 2, Appli
3	90	11.5	852	US-10-170-205E-3413	Sequence 3413, Ap
4	89	11.4	1627	US-10-170-205E-16659	Sequence 16659, A
5	89	11.4	1905	PCT-US04-09388-9	Sequence 9, Appli
6	87	11.2	461	US-60-556-841-11622	Sequence 11622, A
7	86	11.0	841	US-60-565-632-7906	Sequence 7906, Ap
8	86	11.0	841	US-60-579-062-7906	Sequence 7906, Ap
9	85	10.9	295	US-10-425-115-312468	Sequence 312468, A
10	84	10.8	892	US-09-952-267B-5	Sequence 5, Appli
11	84	10.8	892	US-10-872-769-5	Sequence 5, Appli
12	84	10.8	892	US-10-872-769-5	Sequence 5, Appli
13	84	10.8	1507	US-60-579-902-6263	Sequence 6263, Ap
14	83.5	10.7	415	US-09-490-324-280	Sequence 280, App
15	83.5	10.7	873	US-09-952-267B-13	Sequence 13, Appl
16	83.5	10.7	873	US-10-872-769-13	Sequence 13, Appl
17	83.5	10.7	873	US-10-872-769-13	Sequence 13, Appl
18	83	10.6	358	US-09-248-796A-22578	Sequence 22578, A
19	83	10.6	556	US-10-425-115-337674	Sequence 337674, A
20	83	10.6	573	US-60-565-632-7907	Sequence 7907, Ap
21	83	10.6	573	US-60-579-062-7907	Sequence 7907, Ap
22	82.5	10.6	434	US-10-045-674A-594	Sequence 594, App
23	81.5	10.4	177	US-10-737-290-171	Sequence 171, App
24	81.5	10.4	400	US-10-490-953-13	Sequence 13, Appl
25	81.5	10.4	400	US-10-490-953-14	Sequence 14, Appl
26	81.5	10.4	400	US-10-490-953-20	Sequence 20, Appl

ALIGNMENTS

RESULT 1

US-09-741-873C-4
; Sequence 4: Application US/09741873C

; GENERAL INFORMATION:

; APPLICANT: Normark, Staffan

; APPLICANT: Olsen, Arne

; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation

; FILE REFERENCE: 012889-084

; CURRENT APPLICATION NUMBER: US/09/741,873C

; CURRENT FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: SE 8801723-1

; PRIOR FILING DATE: 1998-05-06

; PRIOR APPLICATION NUMBER: US 08/978,878

; PRIOR FILING DATE: 1997-11-26

; PRIOR APPLICATION NUMBER: US 07/347,189

; PRIOR FILING DATE: 1989-05-04

; PRIOR APPLICATION NUMBER: US 07/789,437

; PRIOR FILING DATE: 1994-11-06

; PRIOR APPLICATION NUMBER: US 07/970,846

; PRIOR FILING DATE: 1992-11-03

; PRIOR APPLICATION NUMBER: US 08/187,865

; PRIOR FILING DATE: 1994-01-28

; PRIOR APPLICATION NUMBER: US 08/318,519

; PRIOR FILING DATE: 1994-10-05

; NUMBER OF SEQ ID NOS: 11

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 4

; LENGTH: 151

; TYPE: PRT

; ORGANISM: Escherichia coli

US-09-741-873C-4

Query Match 66.7%; Score 520; DB 5; Length 151;

Best Local Similarity 68.2%; Pred. No. 2.6e-39;

Matches 103; Conservative 21; Mismatches 27; Indels 0; Gaps 0;

Qy	1	MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNSGSPDYDOLVTRVWTHMAHALQ	60
Db	1	MKLLKVAFAAIVFSGSAVAGVVPQWGGGNGGNSGPNSELNIYQYGGNSALALQ	60
Qy	61	SDARKSETTITGSGYNGADVGQADNSTIETQNGFRNNATIDQWAKNSITVGYQYG	120
Db	61	TDARNSDLATITGQGGNGADVGQSDSDSIDLTQRFNGNSATIDQWNGKNSMTVQFGG	120
Qy	121	NNALVNOTASDSSVMVRQVGFNNATANYQ	151
Db	121	NGAAVDQTASNSSVNVTVQGFNNATAHOY	151

RESULT 2
 US-09-741-873C-2
 ; Sequence 2, Application US/09741873C
 ; GENERAL INFORMATION:
 ; APPLICANT: Olses, Arne
 ; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
 ; FILE REFERENCE: 012889-084
 ; CURRENT APPLICATION NUMBER: US/09/741,873C
 ; CURRENT FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: SE 8801723-1
 ; PRIOR FILING DATE: 1998-05-06
 ; PRIOR APPLICATION NUMBER: US 08/978,878
 ; PRIOR FILING DATE: 1997-11-26
 ; PRIOR APPLICATION NUMBER: US 07/347,189
 ; PRIOR FILING DATE: 1989-05-04
 ; PRIOR APPLICATION NUMBER: US 07/789,437
 ; PRIOR FILING DATE: 1991-11-06
 ; PRIOR APPLICATION NUMBER: US 07/970,846
 ; PRIOR FILING DATE: 1992-11-03
 ; PRIOR APPLICATION NUMBER: US 08/187,865
 ; PRIOR FILING DATE: 1994-01-28
 ; PRIOR APPLICATION NUMBER: US 08/318,519
 ; PRIOR FILING DATE: 1994-10-05
 ; NUMBER OF SEQ ID NOS: 11
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 2
 ; LENGTH: 131
 ; TYPE: PRT
 ; ORGANISM: Escherichia coli
 US-09-741-873C-2

Query Match 56.9%; Score 444; DB 5; Length 131;
 Best Local Similarity 65.6%; Pred. No. 1.5e-32;
 Matches 86; Conservative 19; Mismatches 26; Indels 0; Gaps 0;
 QY 21 GVPVQGGGNGHNGGNSGPDYDQVTRVVTHEMAHALQSDARKSETTITQSGYNGAD 80
 DB 1 GVPVQGGGNGHNGGNSGPNSELTQYGGGNSALALQTDARNSDLTITQHGNGAD 60
 QY 81 VCGGADNSTIETONGFRNATIDOWNAKNSDITVGOYGGNNAALVNOTASDSSVMVQV 140
 DB 61 VCGGDDSSIDITQGFNGSATLDOWNKNSMTVQFGGNGAAGVQDTASNSVNVTVQV 120
 QY 141 GFGNNATANQY 151
 DB 121 GFGNNATAHQY 131

RESULT 3
 US-10-170-205E-3413
 ; Sequence 3413, Application US/10170205E
 ; GENERAL INFORMATION:
 ; APPLICANT: ADAMS, Mark
 ; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN
 ; FILE REFERENCE: CL001381
 ; CURRENT APPLICATION NUMBER: US/10/170,205E
 ; CURRENT FILING DATE: 2002-06-13
 ; NUMBER OF SEQ ID NOS: 40312
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 3413
 ; LENGTH: 852
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-170-205E-3413

Query Match 11.5%; Score 90; DB 6; Length 852;
 Best Local Similarity 22.6%; Pred. No. 7.4;
 Matches 45; Conservative 26; Mismatches 56; Indels 72; Gaps 10;
 QY 6 VAAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDYDQL-----VTRVVTHE-----M 55

Db 263 VTGFQLVNYTDTIPAKIMQW-----KNSDARDHTRVDWKRPKYTSALTYDGVKUM 313
 QY 56 AHALQSDARKSETTITQSGYNGADV-----GQAD-----NSTIELTQ 94
 Db 314 AEAFQSLRRQ---RIDISRRGNAGDCLANPAVFWGQIDIQRALQQRFEGLTGNVQFNE 370
 QY 95 NGRFNAT-----DOWNAKNSDI---TVGOYGGNNAALVNOT-----ASD 132
 Db 371 KGRRTNYTLHVEMKHDGIRKGYWNEDDKFVPAATDAAGGDNSSVQNRITVITILED 430
 QY 133 SSVVRQVGFGNNTANQY 151
 Db 431 PYVWLKK-----NANQF 442
 RESULT 4
 US-10-170-205E-16659
 ; Sequence 16659, Application US/10170205E
 ; GENERAL INFORMATION:
 ; APPLICANT: ADAMS, Mark
 ; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN
 ; FILE REFERENCE: CL001381
 ; CURRENT APPLICATION NUMBER: US/10/170,205E
 ; CURRENT FILING DATE: 2002-06-13
 ; NUMBER OF SEQ ID NOS: 40312
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 16659
 ; LENGTH: 1627
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-170-205E-16659
 Query Match 11.4%; Score 89; DB 6; Length 1627;
 Best Local Similarity 24.4%; Pred. No. 19;
 Matches 39; Conservative 18; Mismatches 47; Indels 56; Gaps 7;
 QY 10 AAIIVSGSALAGVV-----PQW-----GGN-----H 32
 Db 621 AAAAKSGHAWGAANQEDKSPTWGPEPKPSQHWGQGRSNPAWSAGGDWADSSVLGH 680
 QY 33 NGGNSGPDYDQVTRVVTHEMAHALQSDARKSETTIT-QSGYNGADVGGADNSTIE 91
 Db 681 LGGKKNKSGWD-----ADNRSGGWDITTRSGNSGWNSTNTKANPGTNWGE 729
 QY 92 LTQGFERNATIDOWNAKNSDITVGOYGGNNAALVNOTAS 131
 Db 730 TLKPGPQON-----WASKPDNNVSNWGG--AASVKQTGT 762

RESULT 5
 PCT-US04-09388-9
 ; Sequence 9, Application PC/TUS0409388
 ; GENERAL INFORMATION:
 ; APPLICANT: INCYTE CORPORATION; MARQUIS, Joseph P.;
 ; APPLICANT: TRAN, Uyen K.; YANG, Yonghong G.;
 ; APPLICANT: TANG, Y. Tom; CHAWLA, Narinder K.;
 ; APPLICANT: MURAGE, Jaji; WANG, Jonathan T.;
 ; APPLICANT: ELLIOTT, Vicki S.; CHIEN, David;
 ; APPLICANT: YUE, Henry; AZIMZAI, Yalda;
 ; APPLICANT: ISON, Craig H.; KHARE, Reena;
 ; APPLICANT: JIN, Pei; RAMKUMAR, Jayalaxmi;
 ; APPLICANT: FAVERO, Kristin D.; RICHARDSON, Thomas W.,
 ; APPLICANT: HAFALIA, April J.A.; BAUGHN, Mariah R.,
 ; APPLICANT: BECHA, Shanya D.; WILSON, Amy D.
 ; TITLE OF INVENTION: PROTEINS ASSOCIATED WITH CELL GROWTH, DIFFERENTIATION, AND DEATH
 ; FILE REFERENCE: PF-1531 PCT
 ; CURRENT APPLICATION NUMBER: PCT/US04/09388
 ; CURRENT FILING DATE: 2004-04-01
 ; PRIOR APPLICATION NUMBER: US 60/457,403
 ; PRIOR FILING DATE: 2003-03-24
 ; PRIOR APPLICATION NUMBER: US 60/465,568
 ; PRIOR FILING DATE: 2003-04-24


```
; PRIOR APPLICATION NUMBER: US 60/476,135
; PRIOR FILING DATE: 2003-06-04
; PRIOR APPLICATION NUMBER: US 60/476,583
; PRIOR FILING DATE: 2003-06-06
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PERL Program
; SEQ ID NO 9
; LENGTH: 1905
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No.: 7525307CD1
PCT-US04-09388-9

Query Match      11.4%; Score 89; DB 1; Length 1905;
Best Local Similarity 24.4%; Pred. No. 23;
Matches 39; Conservative 18; Mismatches 47; Indels 56; Gaps 7;

Qy 10 AAIVVSGSALAGV-----PWG-----GGN-----H 32
Db 759 AAAKSGHAWGAAQEDKSPGWGPPKPKSHWGDGQSRNPAWSAGGWDWADSSVLGH 818
Qy 33 NGGNSSGPDYDQLVTRVVTHEMAHALQSDARKSETTIT-OSGYCNGADVCGGADNSTIE 91
Db 819 LGDGKKNGSGWD-----ADNSRSGSGWNDITRSGNSGWSNSTNTPKANPCTNWGE 867
Qy 92 LTQCFRNATIDOWNNAKNSDITVGYGGNNAALVNQTAS 131
Db 868 TLKPGPQN-----WASRPQDNVNSNWGG--AASVKQIGT 900

RESULT 6
US-60-556-841-11622
; Sequence 11622, Application US/60556841
; GENERAL INFORMATION:
; APPLICANT: Abad Mark S.
; TITLE OF INVENTION: Genes and Uses for Plant Improvement
; FILE REFERENCE: 38-21(53450)
; CURRENT APPLICATION NUMBER: US/60/556,841
; CURRENT FILING DATE: 2004-03-25
; NUMBER OF SEQ ID NOS: 12463
; SEQ ID NO 11622
; LENGTH: 461
; TYPE: PRT
; ORGANISM: Cytophaga hutchinsonii
US-60-556-841-11622

Query Match      11.2%; Score 87; DB 7; Length 461;
Best Local Similarity 24.2%; Pred. No. 6.7;
Matches 40; Conservative 22; Mismatches 69; Indels 34; Gaps 7;

Qy 4 LKVAAP--AAIVVSGSALAGVVPQWGGGNNHNG--GNSSGPDYDQLVTRVVTHEMAHA 58
Db 177 VVGAFHTMAIKTDGSL-----WAGSNNNGRLGIGITTTTKNAPTGVGTATNWKSVYA 229
Qy 59 LQSD--ARKSETTITQSYGNGADVCGGADNSTIELTQCFRNATIDOWN----- 107
Db 230 AGSNTFALKTDGSLWANGYNVAGVLGDGTTDRLSFVQIGTDNN-----WKTISDGNIS 284
Qy 108 -AKNSDITVGYGGNNAALVNQTASDSSVMVRQVGFGNNTANQY 151
Db 285 FALKSDGLTWVWGDNNNSGQFGNGTTTNSLPIQIG-----TDNQW 324

RESULT 7
US-60-565-632-7906
; Sequence 7906, Application US/60565632
; GENERAL INFORMATION:
; APPLICANT: Monsanto Technology, LLC
; APPLICANT: Baum, James A.
; APPLICANT: Kovalic, David K.
; APPLICANT: Larosa, Thomas J.
```

```
; APPLICANT: Lu, Maolong
; APPLICANT: Munyikwa, Tichifa R. I.
; APPLICANT: Roberts, James K.
; APPLICANT: Wu, Wei
; APPLICANT: Zhang, Bei
; TITLE OF INVENTION: Methods for Genetic Control of Insect Infestations in Plants and
; TITLE OF INVENTION: Compositions Thereof
; FILE REFERENCE: 38-21(53403)B
; CURRENT APPLICATION NUMBER: US/60/565,632
; CURRENT FILING DATE: 2004-04-27
; NUMBER OF SEQ ID NOS: 15449
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 7906
; LENGTH: 841
; TYPE: PRT
; ORGANISM: Diabrotica virgifera
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (810)...(810)
; OTHER INFORMATION: Xaa can be any naturally occurring amino acid
US-60-565-632-7906

Query Match      11.0%; Score 86; DB 7; Length 841;
Best Local Similarity 24.7%; Pred. No. 17;
Matches 40; Conservative 17; Mismatches 47; Indels 58; Gaps 7;

Qy 30 GNING--GGNSSGPDYDQLVTRVVTHEMAHALQSDARK---SETTITQSGYNGA---DV 81
Db 360 GNEGTAAGNNANPD-----VGDAAQVNDNGTAAENNGNDAQAQSDN 403
Qy 82 GQGADNSTIELTONG-----FRNNATIDWN-----AKNSDITVGYGG 120
Db 404 GAAAEENTNADAQNGAAQGTANEAENNAENNAADAQNDAAQENGAENGNADAQGT 463
Qy 121 NNAALVNQT-----ASDSSVMVRQVGFGNNTAN 149
Db 464 DNGAAAEENTGNADPAQGNNDNGAAAEENSGNENGIAEENANAD 505

RESULT 8
US-60-579-062-7906
; Sequence 7906, Application US/60579062
; GENERAL INFORMATION:
; APPLICANT: Baum, James A.
; APPLICANT: Kovalic, David K.
; APPLICANT: Larosa, Thomas J.
; APPLICANT: Lu, Maolong
; APPLICANT: Munyikwa, Tichifa R. I.
; APPLICANT: Roberts, James K.
; APPLICANT: Wu, Wei
; APPLICANT: Zhang, Bei
; TITLE OF INVENTION: Methods for Genetic Control of Insect Infestations in Plants and
; TITLE OF INVENTION: Compositions thereof
; FILE REFERENCE: 38-21 (53403) C
; CURRENT APPLICATION NUMBER: US/60/579,062
; CURRENT FILING DATE: 2004-06-11
; NUMBER OF SEQ ID NOS: 41445
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 7906
; LENGTH: 841
; TYPE: PRT
; ORGANISM: Diabrotica virgifera
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (810)...(810)
; OTHER INFORMATION: Xaa can be any naturally occurring amino acid
US-60-579-062-7906

Query Match      11.0%; Score 86; DB 7; Length 841;
Best Local Similarity 24.7%; Pred. No. 17;
Matches 40; Conservative 17; Mismatches 47; Indels 58; Gaps 7;

Qy 30 GNING--GGNSSGPDYDQLVTRVVTHEMAHALQSDARK---SETTITQSGYNGA---DV 81
```

```

Db      360  GNNGTAAAGNANPD-----VQNDAAQVNDNGTAAENNGNADAQSNNDN 403
QY      82  GQCADNSTIELTQNG-----FRNNATIDOWN-----AKNSDITVGYGG 120
Db      404  GAAAEENTVADQNGAAQCTANEAEANNAVADAQNDAAQANENGAAAEENSGNADAAQGT 463
QY      121  NNAALVNQT-----ASDSSVMVRQVGFNNATAN 149
Db      464  DNGAAAEENTGNADPAQGNNDNGAAAEENSGNENGIAAENNANAD 505

RESULT 9
US-10-425-115-312468
; Sequence 312466, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCES: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 312468
; LENGTH: 295
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_48027C.1.pep
US-10-425-115-312468

Query Match      10.9%; Score 85; DB 6; Length 295;
Best Local Similarity 27.1%; Pred. No. 6;
Matches 38; Conservative 14; Mismatches 44; Indels 44; Gaps 5;

QY      26  WGGCGNHNGGNSGPDYDQLVTRVVVTHEMAHALQSOARKSETTITQSGYENGADVGGGA 85
Db      138  YGGGGYSSGGYSSG-GYAANCYGVGSGSGNYSNASGGYSGS-----DGYGNGAASGGYA 192
QY      86  DNSTIELTQNGFRN-----NATIDQNAKN-----SDITVGYGGNNAAL 125
Db      193  NNLSSGYNNGRYNTIGSSDNGTCGYNGYPNPYCAGNVNTGGSSSSGTLGFGGG----- 248
QY      126  VNQTASDSSVMVRQVGFNN 145
Db      249  -----GFGGN 253

RESULT 10
US-09-952-267B-5
; Sequence 5, Application US/09952267B
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USPA1 AND USPA2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCI:024
; CURRENT APPLICATION NUMBER: US/09/952,267B
; CURRENT FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: US/09/336,447
; PRIOR FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 892
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis

```

```
; CURRENT FILING DATE: 2004-06-21
; PRIOR APPLICATION NUMBER: US/09/336,447
; FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 5
; LENGTH: 892
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-10-872-769-5

Query Match 10.8%; Score 84; DB 6; Length 892;
Best Local Similarity 26.8%; Pred. No. 27;
Matches 41; Conservative 17; Mismatches 55; Indels 40; Gaps 9;

QY 28 GGGNHGGGSS---GPDYDOLVTRVVTHEMAHALQSDARKSETTI---TQSGYNGAD 80
DB 75 GKKNEAKGNVSTVGGDYNEAKGNYSF---VGGSSNTAKEKSTIGGGDINDANGYST 132

QY 81 VQGG-----ADNSTI-----ELTQGFNNATIDQWNAKSDITVQYQY---G 120
DB 133 IGGGYISRAIGDSSTIGGGYNNQATGEKSTVAGGRNN---QATGNNSTVAGGSYNQATG 188

QY 121 NNAALV---NOTASDSSVMVRQVGFNNATAN 149
DB 189 NNSTVAGGSHNQATGGSF---AAGVENKANAN 218

RESULT 13
US-60-579-902-6263
; Sequence 6263, Application US/60579902
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Krasomil-Osterfeld, Karina C.
; APPLICANT: Wu, Wei
; TITLE OF INVENTION: Nucleotide and Amino Acid Sequences from Xenorhabdus
; FILE REFERENCE: 38-21 (53444) A
; CURRENT APPLICATION NUMBER: US/60/579,902
; CURRENT FILING DATE: 2004-06-15
; NUMBER OF SEQ ID NOS: 14985
; SEQ ID NO 6263
; LENGTH: 1507
; TYPE: PRT
; ORGANISM: Xenorhabdus bovienii
US-60-579-902-6263

Query Match 10.8%; Score 84; DB 7; Length 1507;
Best Local Similarity 26.5%; Pred. No. 50;
Matches 36; Conservative 20; Mismatches 56; Indels 24; Gaps 7;

QY 28 GGGN--HNGGN---SSGPDYDOLVTRVVTHEMAHALQSDARKSETTITQSGYNGADV 82
DB 961 GKNLVQNGGGTTLITGGNNYTG-----TTTQKTLRQGAFAFSTVSSYTIQNGTLD 1015

QY 83 QGADNSTIELTQNGFR-----NNATIDQWNAKSDITV-GOYGGN--AALVNTASDSS 134
DB 1016 MGGNTTISALSNRVLVGGDNQTVGM-----LTVAGDYSNGNGTVSLSTALADNS 1069

QY 135 VMVRQVGFNNATANQ 150
DB 1070 KTDKLVNGSTSGTTQ 1085

RESULT 14
US-09-490-324-280
; Sequence 280, Application US/09490324
; GENERAL INFORMATION:
; APPLICANT: Knappik, Achim
; APPLICANT: Pack, Peter
; APPLICANT: Ilag, Vic
; APPLICANT: Ga, Liming
; APPLICANT: Moroney, Simon
; TITLE OF INVENTION: USPA1 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/09/952,267B
; CURRENT FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: US/09/336,447
; PRIOR FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98

; TITLE OF INVENTION: Protein/(Poly)peptide libraries
; NUMBER OF SEQUENCES: 373
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: James F. Haley, Jr., Esq. c/o Fish & Neave
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10021
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/490,324
; FILING DATE: 24-Jan-2000
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/025,769
; FILING DATE: 18-FEB-1998
; APPLICATION NUMBER: EP 95 11 3021.0
; FILING DATE: 18-AUG-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: James F. Haley, Jr., Esq.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: MORPHO/5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)596-9000
; TELEFAX: (212)596-9090
; INFORMATION FOR SEQ ID NO: 280:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 415 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 280:
US-09-490-324-280

Query Match 10.7%; Score 83.5; DB 5; Length 415;
Best Local Similarity 29.2%; Pred. No. 12;
Matches 38; Conservative 20; Mismatches 49; Indels 23; Gaps 8;

QY 27 GGG--GNHNGGNSGPF-DYDQVLT---RVVTHEMAHALQSDAR-KSETTITQSGYNG 78
DB 250 GGSSEGGSGGGSGGDFYKRYKANKGAMTENDALQSDAKGLDSVATDYGAAD 309

QY 79 ADVGQAGDNSTIELTQNGFRNNATIDQWNAKSDITVGOYGGNNAALVNQ-----TASDS 133
DB 310 GFIGD-----VSLANG--NGATGDFAGSNQMAQVGS--GDNSPLMNNFRQYLPSPQ 359

QY 134 VMVRQVGF 143
DB 360 SVECRPFVFG 369

RESULT 15
US-09-952-267B-13
; Sequence 13, Application US/09952267B
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USPA1 AND USPA2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/09/952,267B
; CURRENT FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: US/09/336,447
; PRIOR FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
```

```
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 873
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-952-267B-13

Query Match      10.7%; Score 83.5; DB 5; Length 873;
Best Local Similarity 24.8%; Pred. No. 29;
Matches 40; Conservative 20; Mismatches 44; Indels 57; Gaps 10;

QY      4 LKAAAPAAIIVSGSALAGVVPQWGGGNH--GGGNS-----SGPDYDQLVTRVW 51
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
34 LLIIVGILGMATTASAQQTIAQC--GKGMHSIIIGGNDNEANGDYSTVSGGDYNE----- 85
QY      52 THEMAHALQSDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNS 111
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
86 -----AKGDSSTIGGGYNEAN-----GDSSTI---GGGFYN-----EAKGE 119
QY      112 DITVQYCGGNAALVNQRTASDSSVVRQVQFG--NNATANQY 151
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
120 SSTIG--GGDN-----NSATGMVSTIGGGDNNNSATCRY 150
```

Search completed: August 2, 2004, 15:29:52
Job time : 17.8 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:39:53 ; Search time 9.4 Seconds
(without alignments)

1545.204 Million cell updates/sec

Title: US-09-543-407-18

Perfect score: 780

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVGFNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 78:*

1: Pir1.*

2: Pir2.*

3: Pir3.*

4: Pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	691	88.6	151	2 JC6039	fimbrin protein ag
2	691	88.6	151	2 A10635	major curlin chain
3	523	67.1	151	2 S70788	curlin protein cs9
4	501.5	64.3	152	2 D30806	curlin major subun
5	501.5	64.3	152	2 H85665	hypothetical prote
6	114	14.6	1748	2 S42136	cnjB protein - Tet
7	107.5	13.8	151	2 S70787	curlin nucleator p
8	107.5	13.8	151	2 C30806	minor curlin subun
9	107.5	13.8	151	2 G85685	curlin minor chain
10	106.5	13.7	151	2 JC6040	fimbrin protein ag
11	106.5	13.7	151	2 AH0635	nucleation compone
12	95.5	12.2	145	2 AD3143	conserved hypotet
13	95.5	12.2	145	2 H98144	hypothetical prote
14	93.5	12.0	582	2 F70675	probable PPE prote
15	93	11.9	141	2 A33143	hypothetical prote
16	93	11.9	141	2 B38145	hypothetical prote
17	92	11.8	401	2 C88571	protein C05B5.3 [i
18	91.5	11.7	590	1 A45621	leishmanolysin (EC
19	91.5	11.7	2174	2 E39646	hypothetical glyci
20	91	11.7	590	2 E70945	probable PPE prote
21	90	11.5	602	1 PL0221	leishmanolysin (EC
22	90	11.5	896	2 A41273	glutamate receptor
23	90	11.5	906	2 A40222	glutamate receptor
24	90	11.5	906	2 S25852	glutamate receptor
25	89.5	11.5	599	2 B42049	leishmanolysin (EC
26	88.5	11.3	1655	2 E97835	hypothetical prote
27	88	11.3	599	2 A44951	leishmanolysin (EC
28	88	11.3	906	2 S38723	glutamate receptor
29	87.5	11.2	552	2 D70604	Probable PPE prote

30 87.5 11.2 1028 2 A56038 DNA-binding protei
31 87.5 11.2 1213 2 S16356 ovo protein - frui
32 87.5 11.2 1567 2 S11672 ice nucleation pro
33 87 11.2 407 2 T21956 hypothetical prote
34 86 11.0 409 2 T20847 hypothetical prote
35 86 11.0 495 2 B71360 hypothetical prote
36 85.5 11.0 588 2 F70971 hypothetical glyci
37 85 10.9 364 2 S43574 C05B5.3 protein (c
38 85 10.9 645 2 F70825 probable PPE prote
39 85 10.9 3716 2 E70969 probable PPE prote
40 85 10.9 5188 2 B85547 probable RTX fami
41 85 10.9 5291 2 F90696 hypothetical prote
42 84.5 10.8 539 2 C42049 leishmanolysin (EC
43 84 10.8 586 2 T28667 leishmanolysin (EC
44 84 10.8 633 2 A25473 chorion E2 protein
45 84 10.8 646 1 S19916 leishmanolysin (EC

ALIGNMENTS

RESULT 1

JC6039
fimbrin protein agfa precursor - Salmonella enteritidis
C/Species: Salmonella enteritidis
C/Date: 31-Dec-1996 #sequence revision 31-Dec-1996 #text_change 08-Oct-1999
C/Accession: JC6039; PC6015; A44898
R/Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Baner, P.A.; Kay, W.W.
J. Bacteriol. 178, 662-667, 1996
A/Title: Salmonella enteritidis agfBAC operon encoding thin, aggregative fimbriae.
A/Reference number: JC6039; MUID:96146512; PMID:8550497
A/Accession: JC6039
A/Molecule type: DNA
A/Residues: 1-151 <COL>
A/Cross-references: GB:U43280; NID:91184712; PIDN:AAC43599.1; PID:91184714
A/Accession: PC6015
A/Molecule type: protein
A/Residues: 21-52 <CO2>
A/Experimental source: strain 27655-3b
A/Note: the authors translated the codon ACG for residue 44 as Ile
R/Collinson, S.K.; Emody, L.; Muller, K.H.; Trust, T.J.; Kay, W.W.
J. Bacteriol. 173, 4773-4781, 1991
A/Title: Purification and characterization of thin, aggregative fimbriae from Salmonell
A/Reference number: A44898; MUID:91310586; PMID:1677357
A/Contents: 27655
A/Accession: A44898
A/Status: preliminary
A/Molecule type: protein
A/Residues: 21-33 <CO3>
A/Note: sequence extracted from NCBI backbone (NCBI:45936)
C/Genetics:
A/Gene: agfa
C/Function:
A/Description: major component of thin aggregative fimbriae
A/Note: fimbriae bind to fibronectin, plasminogen, tissue plasminogen activator
C/Keywords: fimbrin
F;1-20/Domain: signal sequence #status predicted <SIG>
F;21-151/Product: fimbrin protein agfa #status experimental <MAT>

Query Match 88.6%; Score 691; DB 2; Length 151;
Best Local Similarity 90.7%; Pred. No. 9.7e-53;
Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSSGPDYDQLVTRVVTHEMAHALQ 60

Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSSGPDSTLSIYGYGSANAALALQ 60

QY 61 SPARKSETTITSGYNGADVGCGADNSTIETQNGFRNNATIDOWNAKNSDITVQYGG 120

Db 61 SPARKSETTITSGYNGADVGCGADNSTIETQNGFRNNATIDOWNAKNSDITVQYGG 120

QY 121 NNAAALVNOTASDSSVMVRQVGFNNATANQY 151

Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 2
AI0635
major curlin chain precursor [imported] - Salmonella enterica subsp. enterica serovar Typhimurium
C/Species: Salmonella enterica subsp. enterica serovar Typhi
A/Note: this species has also been called Salmonella typhi
C/Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C/Accession: AI0635
R/Farkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Mouton, S.; O'Garra, P.
Nature 413, 848-852, 2001
A/Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.
A/Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhimurium
A/Reference number: AB0502; MUID:21534947; PMID:11677608
A/Accession: AI0635
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-151 <PAR>
A/Cross-references: GB:AL513382; PIDN:CAD08268.1; PID:g16502315; GSPDB:GN00176
C/Genetics:
A/Gene: STY1181

Query Match 88.6%; Score 691; DB 2; Length 151;
Best Local Similarity 90.7%; Pred. No. 9.7e-53;
Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIVSGSALAGVVPQGGNGHNGGNSGPDYDQLVTRVVTHEMAHALQ 60
Db 1 MKLLKVAAPAAIVSGSALAGVVPQGGNGHNGGNSGPDYDQLVTRVVTHEMAHALQ 60

QY 61 SPARKSETTITQSGYNGADVGQADNSTIELTQGFNNATIDQWNAKNSDITVQYGG 120
Db 61 SPARKSETTITQSGYNGADVGQADNSTIELTQGFNNATIDQWNAKNSDITVQYGG 120

QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 3
S70788
curlin protein csgA precursor - Escherichia coli (strain K-12)
N/Alternate names: csgA protein; major curlin protein
C/Species: Escherichia coli
C/Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 01-Mar-2002
C/Accession: S70788; G64846; S31202; S34560; S34559
R/Hammar, M.; Arngvist, A.; Bian, Z.; Olsen, A.; Normark, S.
Mol. Microbiol. 18, 661-670, 1995
A/Title: Expression of two csg operons is required for production of fibronectin- and curli
A/Reference number: S70788; MUID:96414468; PMID:8817489
A/Accession: S70788
A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-151 <HAM>
A/Cross-references: EMBL:X90754; NID:g1147558; PIDN:CA62282.1; PID:g1147564
A/Experimental source: strain K12, substrain W3110
A/Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995
R/Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, D.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A/Title: The complete genome sequence of Escherichia coli K-12.
A/Reference number: A64720; MUID:97426617; PMID:9278503
A/Accession: G64846
A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-151 <BLAT>
A/Cross-references: GB:AE000205; GB:U00096; NID:g1787265; PIDN:AAC74126.1; PID:g1787279;
A/Experimental source: strain K-12, substrain MG1655
R/Olsen, A.; Arngvist, A.; Hammar, M.; Sukupolvi, S.; Normark, S.
Mol. Microbiol. 7, 523-536, 1993

A/Title: The RpoS sigma factor relieves H-NS-mediated transcriptional repression of csgA
A/Reference number: S31202; MUID:93211294; PMID:8459772
A/Accession: S31202
A/Molecule type: DNA
A/Residues: 1-6,'V',8-151 <OLG1>
A/Cross-references: EMBL:L04979
A/Accession: S34560
A/Molecule type: protein
A/Residues: 21-42;44-50 <OLS2>
R/Olsen, A.N.; Arngvist, A.M.
submitted to the EMBL Data Library, October 1992
A/Reference number: S34559
A/Accession: S34559
A/Molecule type: DNA
A/Residues: 1-133,'RQDSGLW' <OLS3>
A/Cross-references: EMBL:L04979; NID:g250424; PIDN:AAA23616.1; PID:g290425
A/Experimental source: strain K-12, substrain W3110
C/Genetics:
A/Gene: csgA
A/Map position: 23.15
C/Function:
A/Description: major component of wild-type curli; interaction between CsgA and CsgB tri
A/Note: curli are thin, coiled fibers expressed on the surface of Escherichia coli that
and H-kininogen; in the absence of CsgA, CsgB can self-assemble into polymers
F/1-20/Domain: signal sequence #status predicted <SIG>
F/2-151/Product: curlin #status experimental <MAT>

Query Match 67.1%; Score 523; DB 2; Length 151;
Best Local Similarity 68.9%; Pred. No. 3.1e-38;
Matches 104; Conservative 20; Mismatches 27; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIVSGSALAGVVPQGGNGHNGGNSGPDYDQLVTRVVTHEMAHALQ 60
Db 1 MKLLKVAAPAAIVSGSALAGVVPQGGNGHNGGNSGPDYDQLVTRVVTHEMAHALQ 60

QY 61 SPARKSETTITQSGYNGADVGQADNSTIELTQGFNNATIDQWNAKNSDITVQYGG 120
Db 61 SPARKSETTITQSGYNGADVGQADNSTIELTQGFNNATIDQWNAKNSDITVQYGG 120

QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 4
D90806
curlin major subunit CsgA [imported] - Escherichia coli (strain O157:H7, substrain RIMD
C/Species: Escherichia coli
C/Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C/Accession: D90806
R/Hayaishi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
Gawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.
DNA Res. 8, 11-22, 2001
A/Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen
A/Reference number: A99629; MUID:21156231; PMID:11258796
A/Accession: D90806
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-152 <HAY>
A/Cross-references: GB:BA000007; PIDN:BA034843.1; PID:g13360880; GSPDB:GN00154
A/Experimental source: strain O157:H7, substrain RIMD 0509952
C/Genetics:
A/Gene: ECs1420

Query Match 64.3%; Score 501.5; DB 2; Length 152;
Best Local Similarity 67.1%; Pred. No. 2.3e-36;
Matches 102; Conservative 21; Mismatches 28; Indels 1; Gaps 1;

QY 1 MKLLKVAAPAAIVSGSALAGVVPQGGNGHNGGNSGPDYDQLVTRVVTHEMAHALQ 59
Db 1 MKLLKVAAPAAIVSGSALAGVVPQGGNGHNGGNSGPDYDQLVTRVVTHEMAHALQ 60

QY 60 QSDARKSETTITQSGYNGADVGQADNSTIELTQGFNNATIDQWNAKNSDITVQYGG 119

Db 61 QADARNSDLTTHQGGNGADVGCGSDSDSDITDQRGFGNSATLDQWNGKDSHMTVKQFG 120

QY 120 GNNAAALVNQATSDSSVMVRQVGFNNATANQY 151

Db 121 GGNGAAVDQTASNSTVNVTVQVGFNNATAHQY 152

RESULT 5

H85665

hypothetical protein csgA [imported] - Escherichia coli (strain O157:H7, substrain ED933)

C:Species: Escherichia coli

C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001

C:Accession: H85665

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew, L.; Flier, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouisis, K.; Apodaca, Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: H85665

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-152 <STO>

A:Cross-references: GB:AE0051174; NID:G12514574; PIDN:AAG55788.1; GSPDB:GN00145; UNGP:216

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: csgA

Query Match 64.3%; Score 501.5; DB 2; Length 152;

Best Local Similarity 67.1%; Pred. No. 2.3e-36;

Matches 102; Conservative 21; Mismatches 28; Indels 1; Gaps 1;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQW-GCGGNHNGGNSGSPDYDQVLTWVTHMAHAL 59

Db 1 MKLLKVAALAAIVFSGSALAGVVPQYGGGGNHCGGGNSGPNSELNIYQVGGNSALAL 60

QY 60 QSDARKSETTITQSGYNGADVGGAGDNSTIETLQNGFRNNATIDQWNAKNSDITVQYQG 119

Db 61 QADARNSDLTTHQGGNGADVGCGSDSDSDITDQRGFGNSATLDQWNGKDSHMTVKQFG 120

QY 120 GNNAAALVNQATSDSSVMVRQVGFNNATANQY 151

Db 121 GGNGAAVDQTASNSTVNVTVQVGFNNATAHQY 152

RESULT 6

S42136

cnjB protein - Tetrahymena thermophila

C:Species: Tetrahymena thermophila

C:Date: 19-Mar-1997 #sequence_revision 18-Jul-1997 #text_change 07-Dec-1999

C:Accession: S42136; S42135; S03650

R:Taylor, F.M.; Martindale, D.W.

submitted to the EMBL Data Library, October 1992

A:Reference number: S42136

A:Accession: S42136

A:Molecule type: DNA

A:Residues: 1-1748 <TAY>

A:Cross-references: EMBL:L03710; NID:g161751; PID:g161752

R:Taylor, F.M.; Martindale, D.W.

Nucleic Acids Res. 21, 4610-4614, 1993

A:Title: Retroviral-type zinc fingers and glycine-rich repeats in a protein encoded by c

A:Reference number: S42135; MUID:94051569; PMID:8233798

A:Accession: S42135

A:Molecule type: DNA

A:Residues: 1164-1174; 1179-1198; 1233-1252; 1285-1293; 1297-1309; 1316-1326; 1331-1341; 1343-1

A:Cross-references: EMBL:L03710

R:Martindale, D.W.; Taylor, F.M.

Nucleic Acids Res. 15, 2189-2201, 1988

A:Title: Multiple introns in a conjugation-specific gene from Tetrahymena thermophila.

A:Reference number: S03650; MUID:8619811; PMID:335771

A:Accession: S03650

A:Molecule type: DNA

A:Residues: 236-250, 'I', 252-255, 'N', 257-773 <NAR>

Best Local Similarity 35.4%; Pred. No. 0.024; Matches 29; Conservative 9; Mismatches 43; Indels 1; Gaps 1;

QY 70 ITQSGYNGADVQGGADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGGNNAALVNQT 129
 Db 49 IQGAGTNSAQLRQGGSKLLAVVAQEGSSNRKIDTQTDYNL-AVIDQAGSANDASISQG 107

QY 130 ASDSSVMVRQVGFNNATANQY 151
 Db 108 AYGNTAMIIQKSGNKANITQY 129

RESULT 8
 minor curlin subunit precursor CsgB [imported] - Escherichia coli (strain O157:H7, subsp. C; Species: Escherichia coli
 C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
 C;Accession: G90806
 R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G. Gasawara, N.; Yasunaga, T.; Kuhara, T.; Shiba, T.; Hattori, M.; Shinagawa, H. DNA Res. 8, 11-22, 2001
 A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic islands
 A;Reference number: A95629; MUID:21156231; PMID:11258796
 A;Accession: G90806
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-151 <STO>
 A;Cross-references: GB:BA000007; PIDN:BA034842.1; PID:gl3360879; GSPDB:GN00154
 A;Experimental source: strain O157:H7, substrain RIMD 0509952
 C;Genetics:
 A;Gene: Ecs1419

Query Match 13.8%; Score 107.5; DB 2; Length 151;
 Best Local Similarity 35.4%; Pred. No. 0.024; Matches 29; Conservative 9; Mismatches 43; Indels 1; Gaps 1;

QY 70 ITQSGYNGADVQGGADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGGNNAALVNQT 129
 Db 49 IQGAGTNSAQLRQGGSKLLAVVAQEGSSNRKIDTQTDYNL-AVIDQAGSANDASISQG 107

QY 130 ASDSSVMVRQVGFNNATANQY 151
 Db 108 AYGNTAMIIQKSGNKANITQY 129

RESULT 9
 curlin minor chain precursor, CsgA homolog [imported] - Escherichia coli (strain O157:H7; C; Species: Escherichia coli
 C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 C;Accession: G85665
 R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca Nature 409, 529-533, 2001
 A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A;Reference number: A85480; MUID:21074935; PMID:11206551
 A;Accession: G85665
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-151 <STO>
 A;Cross-references: GB:AB005174; NID:gl2514573; PIDN:AA055787.1; GSPDB:GN00145; UWGP:216
 A;Experimental source: strain O157:H7, substrain EDL933
 C;Genetics:
 A;Gene: csgB

Query Match 13.8%; Score 107.5; DB 2; Length 151;
 Best Local Similarity 35.4%; Pred. No. 0.024; Matches 29; Conservative 9; Mismatches 43; Indels 1; Gaps 1;

QY 70 ITQSGYNGADVQGGADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGGNNAALVNQT 129
 Db 49 IQGAGTNSAQLRQGGSKLLAVVAQEGSSNRKIDTQTDYNL-AVIDQAGSANDASISQG 107

QY 130 ASDSSVMVRQVGFNNATANQY 151
 Db 108 AYGNTAMIIQKSGNKANITQY 129

RESULT 10
 fimbria protein agfB precursor - Salmonella enteritidis
 C;Species: Salmonella enteritidis
 C;Date: 31-Dec-1996 #sequence_revision 31-Dec-1996 #text_change 08-Oct-1999
 C;Accession: JC6040
 R;Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Bansen, P.A.; Kay, W.W. J. Bacteriol. 178, 662-667, 1996
 A;Title: Salmonella enteritidis agfBAC operon encoding thin, aggregative fimbriae.
 A;Reference number: JC6039; MUID:96146512; PMID:8550497
 A;Accession: JC6040
 A;Molecule type: DNA
 A;Residues: 1-151 <COL>
 A;Cross-references: GB:U43280; NID:gl184712; PIDN:AA043598.1; PID:gl184713
 A;Experimental source: strain 276755-3b
 C;Genetics:
 A;Gene: agfB
 A;Function:
 A;Description: minor component of thin aggregative fimbriae
 A;Note: fimbriae bind to fibronectin, plasminogen, tissue plasminogen activator
 C;Keywords: fimbria
 F;1-21/Domain: signal sequence #status predicted <SIG>
 F;22-151/Product: fimbria protein agfB #status predicted <MAT>

Query Match 13.7%; Score 106.5; DB 2; Length 151;
 Best Local Similarity 30.7%; Pred. No. 0.029; Matches 35; Conservative 17; Mismatches 43; Indels 19; Gaps 5;

QY 55 MAHALQSDARKSE-----TTITQSGYNGADVQGGADNST-----IELTQNGFR 98
 Db 18 IATATNYDLARSEYNFAVNELSKSFNQAAIIIGVGTDSARVRQEGSKLLSVISQEGN 77

QY 99 NNATIDQWNAKNSDIT-VQYGGNNAALVNQTASDSSVMVRQVGFNNATANQY 151
 Db 78 NRAKVDQ--AGNYNFAYIEQTGNANDASISQAYGNSAAIIQKSGNKANITQY 129

RESULT 11
 AH0635
 nucleation component of curlin monomers [imported] - Salmonella enterica subsp. enterica
 C;Species: Salmonella enterica subsp. enterica serovar Typhi
 A;Note: this species has also been called Salmonella typhi
 C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
 C;Accession: AH0635
 R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, Th. T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P. Nature 413, 848-852, 2001
 A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K. A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Paratyphi A
 A;Reference number: AB0502; MUID:21534947; PMID:11677608
 A;Accession: AH0635
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-151 <PAR>
 A;Cross-references: GB:AL513382; PIDN:CAD08267.1; PID:gl65023.4; GSPDB:GN00176
 C;Genetics:
 A;Gene: STY1180

Query Match 13.7%; Score 106.5; DB 2; Length 151;
 Best Local Similarity 30.7%; Pred. No. 0.029; Matches 35; Conservative 17; Mismatches 43; Indels 19; Gaps 5;

QY 55 MAHALQSDARKSE-----TTITQSGYNGADVQGGADNST-----IELTQNGFR 98
 Db 18 IATATNYDLARSEYNFAVNELSKSFNQAAIIIGVGTDSARVRQEGSKLLSVISQEGN 77

QY 99 NNATIDQWNAKNSDIT-VQYGGNNAALVNQTASDSSVMVRQVGFNNATANQY 151

DB 78 NRAKVQ--AGNYNFAYIEOTGNANDASISQSAYGNSAAIIQKSGNKANITYQ 129

RESULT 12

AD3143

Conserved hypothetical protein Atu4768 [imported] - Agrobacterium tumefaciens (strain C58)

C;Species: Agrobacterium tumefaciens

C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002

C;Accession: AD3143

R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.; Erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutuyavin, T.; Levy, R.; Li, M.; McClellan, R.; Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001

A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, J.E.

A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A;Reference number: AB2577; MUID:21608550; PMID:11743193

A;Accession: AD3143

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-145 <KUR>

A;Cross-references: GB:AE086899; PIDN:AAU45562.1; PID:gl7743277; GSPDB:GNO0187

A;Experimental source: strain C58 (Dupont)

C;Genetics:

A;Gene: Atu4768

A;Map position: linear chromosome

Query Match 12.2%; Score 95.5; DB 2; Length 145;

Best Local Similarity 24.2%; Pred. No. 0.25; 57; Indels 43; Gaps 4;

Matches 39; Conservative 22; Mismatches 57; Indels 43; Gaps 4;

QY 3 LLKVAFAAIVVSGSALAGVVP-----QWG-----GGNHNGGNSGGPDYDQLVTR 49

DB 1 MIRKSFTASALVALVGLSAAAPAMANDVRIEQYGSNSAGAGQEGYGNRIRTYQNGGYNR 60

QY 50 VVTHEMAHALQSDARKSETTITQSGYNGADVGQADNSTIELTONGFRNATIDQWNAK 109

DB 61 IVGH-----QYGRHLNSAVGQEGHDNYGSTTQNGRNVAGI----- 96

QY 110 NSDITVQYGGNNAALYNQTASDSSVMVRQVGFGNNAATQ 150

DB 97 -----QFGSGNHTTILTDGNGNIAAGVQVGRGCSANVSQ 131

RESULT 13

H98144

hypothetical protein AGR_L_228 [imported] - Agrobacterium tumefaciens (strain C58, Cerec)

C;Species: Agrobacterium tumefaciens

C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002

C;Accession: H98144

R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.

Science 294, 2323-2328, 2001

A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens C58

A;Reference number: A97359; MUID:21608551; PMID:11743194

A;Accession: H98144

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-145 <KUR>

A;Cross-references: GB:AE007870; PIDN:AAK8682.1; PID:gl5158413; GSPDB:GNO0170

C;Genetics:

A;Gene: AGR_L_228

A;Map position: linear chromosome

Query Match 12.2%; Score 95.5; DB 2; Length 145;

Best Local Similarity 24.2%; Pred. No. 0.25; 57; Indels 43; Gaps 4;

Matches 39; Conservative 22; Mismatches 57; Indels 43; Gaps 4;

QY 3 LLKVAFAAIVVSGSALAGVVP-----QWG-----GGNHNGGNSGGPDYDQLVTR 49

DB 1 MIRKSFTASALVALVGLSAAAPAMANDVRIEQYGSNSAGAGQEGYGNRIRTYQNGGYNR 60

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:36:12 ; Search time 5.3 Seconds
(without alignments)

1483.508 Million cell updates/sec

Title: US-09-543-407-18

Perfect score: 730

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMYRVQVFGNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	691	88.6	151	1	CSGA_SALTY
2	523	67.1	151	1	CSGA_ECOLI
3	501.5	64.3	152	1	CSGA_ECO57
4	107.5	13.8	151	1	CSGB_ECOLI
5	106.5	13.7	151	1	CSGB_SALTY
6	106.5	13.7	151	1	CSGB_SALTY
7	92	11.8	401	1	YK03_CAREL
8	92	11.8	1656	1	OMP6_RICGA
9	91.5	11.7	590	1	GP63_LEIDO
10	91	11.7	1093	1	PER_DROWI
11	90	11.5	602	1	GP63_LEIMA
12	90	11.5	906	1	GLR1_HUMAN
13	88.5	11.3	1655	1	OMP6_RICCN
14	88	11.3	599	1	GP63_LEICH
15	87.5	11.2	1028	1	OVN_DROME
16	87.5	11.2	1567	1	ICEN_XANTC
17	86	11.0	485	1	Y136_TREPA
18	86	11.0	493	1	GATA_RHIME
19	84	10.8	646	1	GP63_LEIME
20	84	10.8	907	1	GLR1_MOUSE
21	84	10.8	907	1	GLR1_RAT
22	83	10.6	720	1	KKE6_YEAST
23	83	10.6	947	1	SECA_STRCO
24	82	10.5	342	1	OMP6_RAHQ
25	81.5	10.4	424	1	COAA_BPFD
26	81.5	10.4	424	1	COAA_BPMD3
27	81.5	10.4	493	1	GATA_AGRF5
28	81	10.4	365	1	ROAL_DROME
29	81	10.4	576	1	DEAF_DROME
30	81	10.4	678	1	YF48_MYCTU
31	81	10.4	1034	1	ICEN_PANAN
32	80.5	10.3	641	1	IND_ARTGO
33	80.5	10.3	1185	1	MAPX_DROME

ALIGNMENTS

RESULT 1	CSGA_SALTY	STANDARD;	PRT;	151 AA.
ID	CSGA_SALTY	STANDARD;	PRT;	151 AA.
AC	P55225			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DE	Major curlin subunit precursor (Fimbrin SEF17)			
GN	CSGA OR AGFA OR STM1144 OR STY1181 OR T1776			
OS	Salmonella typhimurium			
OS	Salmonella typhi, and			
OS	Salmonella enteritidis			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OC	Enterobacteriaceae; Salmonella			
OX	NCBI_TaxID=602, 601, 592;			
RN	[1] SEQUENCE FROM N.A.			
RP	SPECIES=Salmonella typhimurium; STRAIN=SR-11;			
RC	MEDLINE=98117058; PubMed=9457880;			
RX	Romling U., Bian Z., Hammar M., Sierralta W.D., Normark S.;			
RA	"Curli fibers are highly conserved between Salmonella typhimurium and			
RT	Escherichia coli with respect to operon structure and regulation.";			
RL	J. Bacteriol. 180:722-731(1998).			
RN	[2] SEQUENCE FROM N.A.			
RP	SPECIES=Salmonella typhimurium; STRAIN=IT2 / SGSC1412 / ATCC 700720;			
RC	MEDLINE=21534948; PubMed=11677609;			
RX	McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,			
RA	Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,			
RA	Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,			
RA	Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,			
RA	Waterston R., Wilson R.K.;			
RT	"Complete genome sequence of Salmonella enterica serovar Typhimurium			
RL	IT2.";			
RN	Nature 413:852-856 (2001).			
RN	[3] SEQUENCE FROM N.A.			
RP	SPECIES=Salmonella typhi; STRAIN=CT18;			
RC	MEDLINE=21534947; PubMed=11677608;			
RX	Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,			
RA	Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebatia M.,			
RA	Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,			
RA	Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,			
RA	Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,			
RA	Krogh A., Jansen T.S., Leather S., Moule S., O'Gaora P., Parry C.,			
RA	Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,			
RA	Whitehead S., Barrrell B.G.;			
RT	"Complete genome sequence of a multiple drug resistant Salmonella			
RL	enterica serovar Typhi CT18.";			
RN	Nature 413:848-852(2001).			
RN	[4] SEQUENCE FROM N.A.			
RP	SPECIES=Salmonella typhi; STRAIN=Ty2 / ATCC 700931;			
RC	MEDLINE=22531367; PubMed=12644504;			
RX	Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,			
RA	Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;			

Q03646 plasmodium
P95759 streptomyce
P14914 rickettsia
Q53047 r outer mem
P26968 tenebrio mo
Q04893 saccharomyc
P57037 neisseria m
P24785 drosophila
P75139 mycoplasma
P43096 candida alb
P16239 erwinia her
P33666 escherichia

```

RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RL and CT18.";
RN [5]
RP SEQUENCE FROM N.A.
RC SPECIES=S.enteritidis; STRAIN=27655-3B;
RX MEDLINE=96146512; PubMed=8550497;
RA Collinson S.K., Clouthier S.C., Doran J.L., Baner P.A., Kay W.W.;
RT "Salmonella enteritidis agfBAC operon encoding thin, aggregative
fimbriae.";
RN [6]
RL J. Bacteriol. 178:662-667(1996).
RP SEQUENCE OF 21-151 FROM N.A.
RC SPECIES=S.enteritidis; STRAIN=27655-3B;
RX MEDLINE=94013373; PubMed=8104955;
RA Doran J.L., Collinson S.K., Burian J., Sarlos G., Todd E.C.D.,
RX Munro C.K., Kay C.M., Baner P.A., Peterkin P.I., Kay W.W.;
RT "DNA-based diagnostic tests for Salmonella species targeting agfA,
the structural gene for thin, aggregative fimbriae.";
RN [7]
RL J. Clin. Microbiol. 31:2263-2273(1993).
RP SEQUENCE OF 21-33.
RC SPECIES=S.enteritidis; STRAIN=27655-3B;
RX MEDLINE=91310586; PubMed=1677357;
RA Collinson S.K., Emeoey L., Mueller K.-M., Trust T.J., Kay W.W.;
RT "Purification and characterization of thin, aggregative fimbriae from
Salmonella enteritidis.";
RN [8]
RL J. Bacteriol. 173:4773-4781(1991).
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
FIBRONECTIN.
CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
DR EMBL; AJ002301; CAA05317.1; -
DR EMBL; AF008749; AAL20074.1; -
DR EMBL; AL627269; CAD08268.1; -
DR EMBL; AF016840; AAO69399.1; -
DR EMBL; U43280; AAC43599.1; -
DR PIR; JC6039; JC6039.
DR StyGene; SG10608; CSGA.
KW Fimbria; Signal; Complete proteome.
FT SIGNAL 1 20
FT CHAIN 21 151 MAJOR CURLIN SUBUNIT.
FT CONFLICT 134 151 SYMPVQGVFGNNATANQY -> DSYTQVAS (IN
REF. 6).
FT SEQUENCE 151 AA; 15305 MW; B7DAC0D16B621359 CRC64;
Query Match 88.6%; Score 691; DB 1; Length 151;
Best Local Similarity 90.7%; Pred. No. 5.3e-53;
Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGGPDYQVLVTVRVVTHMAHALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGGPDYQVLVTVRVVTHMAHALQ 60
QY 61 SDARKSETTITQSGYNGADVCGQADNSTIELTQNGFPNNATIDQWNAKNSDITVGYGG 120
Db 61 SDARKSETTITQSGYNGADVCGQADNSTIELTQNGFPNNATIDQWNAKNSDITVGYGG 120
QY 121 NNAALVNOTASDSSVWVQVQVFGNNATANQY 151
Db 121 NNAALVNOTASDSSVWVQVQVFGNNATANQY 151

```

```

RESULT 2
CSGA_ECOLI
ID CSGA_ECOLI STANDARD; PRT; 151 AA.
AC P28307;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Major curlin subunit precursor.
GN CSGA OR B1042.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / W3110;
RX MEDLINE=93211294; PubMed=8459772;
RA Olsen A., Arngvist A.;
RT "The RpoS sigma factor relieves H-NS-mediated transcriptional
repression of csGA, the subunit gene of fibronectin-binding curli in
Escherichia coli";
RN [2]
RL Mol. Microbiol. 7:523-536(1993).
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MC4100;
RX MEDLINE=96414468; PubMed=8817489;
RA Hammar M., Arngvist A., Bian Z., Olsen A., Normark S.;
RT "Expression of two csG operons is required for production of
fibronectin- and congo red-binding curli polymers in Escherichia coli
K-12.";
RN [3]
RL Mol. Microbiol. 18:661-670(1995).
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RN [4]
RL Science 277:1453-1474(1997).
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itch T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RT "A 718-Kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 12.7-28.0 min region on the linkage map.";
RN [5]
RL DNA Res. 3:137-155(1996).
RP SEQUENCE OF 21-40.
RC STRAIN=K12 / YNML;
RX MEDLINE=93023873; PubMed=1357528;
RA Arngvist A., Olsen A., Pfeifer J., Russell D.G., Normark S.;
RT "The Crl protein activates cryptic genes for curli formation and
fibronectin binding in Escherichia coli HB101.";
RN [6]
RL Mol. Microbiol. 6:2443-2452(1992).
RP SEQUENCE OF 21-31.
RC MEDLINE=91310586; PubMed=1677357;
RA Collinson S.K., Emeoey L., Trust T.J., Kay W.W.;
RT "Purification and characterization of thin, aggregative fimbriae from
Salmonella enteritidis.";
RN [7]
RL J. Bacteriol. 173:4773-4781(1991).
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
FIBRONECTIN.
CC

```



```

RT K-12."
RL Mol. Microbiol. 18:661-670(1995).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=K12 / MGL655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1443-1474 (1997).
RN [3]
RN SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitegawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map."
RL DNA Res. 3:137-155(1996).
RN [4]
RN SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.F., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
RL Nature 409:529-533(2001).
RN [5]
RN SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayaishi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12."
RL DNA Res. 8:11-22(2001).
RN [6]
RN SEQUENCE OF 1-21 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=95157246; PubMed=7854117;
RA Arqvist A., Olsen A., Normark S.;
RT "Sigma S-dependent growth-phase induction of the esgB promoter in
RT Escherichia coli can be achieved in vivo by sigma 70 in the absence
RT of the nucleoid-associated protein H-NS."
RL Mol. Microbiol. 13:1021-1032(1994).
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
CC CURLIN MONOMERS.
CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/)
CC or send an email to license@isb-sib.ch)
CC -----
CC EMBL, X90754; CAA62281.1; -.

```

```

DR EMBL; AE000205; AAC74125.1; -.
DR EMBL; D90741; BAA35831.1; -.
DR EMBL; AE005315; AAG55787.1; -.
DR EMBL; AP002554; BAB34842.1; -.
DR FIR; C90806; C90806.
DR FIR; G85665; G85665.
DR FIR; S70787; S70787.
DR EcoGene; EG12621; csbB.
KW Fimbria; Signal; Complete proteome.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 151 MINOR CURLIN SUBUNIT.
SQ SEQUENCE 151 AA; 15882 MW; B18D266B964014B8 CRC64;

Query Match 13.8%; Score 107.5; DB 1; Length 151;
Best Local Similarity 35.4%; Pred. No. 0.012;
Matches 29; Conservative

QY 70 ITQSGYNGADVGGADNNTIELTONGFRNATIDWNKNSDITVGYQGNRAALVNQT 129
DB 49 IGQAGTNSAQLRQGGSKLLAVVAQEGSSNRAKIDQTGYNL-AYIDQASANDASISQ 107
QY 130 ASDSSVMVRQVGFNNATANOY 151
DB 108 AYGNTAMIIQKSGGNKANITQY 129

RESULT 5
CSGB SALTI STANDARD; PRT; 151 AA.
AC Q87M3;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Minor curlin subunit precursor.
GN CSGB OR STY1180 OR T1777.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davies P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrell B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18."
RL Nature 413:848-852(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=TY2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kocoyanni V., Schwartz D.C., Blattner F.R.;
RT "Comparative Genomics of Salmonella enterica serovar Typhi strains Ty2
RT and CT18."
RL J. Bacteriol. 185:2330-2337(2003).
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
CC CURLIN MONOMERS.
CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

CC EMBL; AL627289; CAD08267.1; -
CC EMBL; AE016840; AAO69400.1; -
CC Fimbrin; Signal; Complete proteome.
CC SIGNAL 1 21 POTENTIAL.
CC FT CHAIN 22 151 MINOR CURLIN SUBUNIT.
CC SQ SEQUENCE 151 AA; 16254 MW; 161C54326E573495 CRC64;

Query Match 13.7%; Score 106.5; DB 1; Length 151;
Best Local Similarity 30.7%; Pred. No. 0.015;
Matches 35; Conservative 17; Mismatches 43; Indels 19; Gaps 5;

QY 55 MAHALQSDARKSE-----TTITQSGYNGADVQ--GADNST-----IELTQNGFR 98
DB 18 IATATNYDLARSEYFVAVNELSKSFNCAAIIGVGTDSARVQRQSGKLLSVISQGEN 77
QY 99 NNATIDQWAKNSDIT-VGYGGNNALVNOTASDSSVMVRQVGFNNATANQY 151
DB 78 NRAKVQD--AGNYNFAYIEQTGNANDASISQAYGNSAALIQKSGNKANITQY 129

RESULT 6

ID CSGB SALTY STANDARD; PRT; 151 AA.
AC P55226;
DT 01-OCT-1996 (Rel. 34; Created)
DT 01-OCT-1996 (Rel. 34; Last sequence update)
DT 28-FEB-2003 (Rel. 41; Last annotation update)
DE Minor curlin subunit precursor (Fimbrin SEF17 minor subunit).
GN CSGB OR AGFB OR STM143.
OS Salmonella typhimurium, and
OS Salmonella enteritidis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602, 592;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=SR-11;
RX MEDLINE=98117058; PubMed=9457880;
RA Romling U., Bian Z., Hammar M., Sierralta W.D., Normark S.;
RT "Curli fibers are highly conserved between Salmonella typhimurium and
RT Escherichia coli with respect to operon structure and regulation.";
RL J. Bacteriol. 180:722-731(1998).
RN [2]

RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC11412 / ATCC 700720;
RX MEDLINE=21514948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2";
RL Nature 413:852-856(2001).
RN [3]

RP SEQUENCE FROM N.A.
RC SPECIES=S.enteritidis; STRAIN=27655-3B;
RX MEDLINE=96146512; PubMed=8550497;
RA Collinson S.K., Clouthier S.C., Doran J.L., Baner P.A., Kay W.W.;
RT "Salmonella enteritidis agfBAC operon encoding thin, aggregative
RT fimbriae";
RL J. Bacteriol. 178:662-667(1996).
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
CC CURLIN MONOMERS.
CC -!- SIMILARITY: BELONGS TO THE CSGB/CSGB FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

CC EMBL; AJ002301; CAA05316.1; -
CC EMBL; AB008749; AAL20073.1; -
CC EMBL; U43280; AAC43598.1; -
CC PIR; JC6040; JC6040.
CC StyGene; SG10609; CSGB.
CC Fimbrin; Signal; Complete proteome.
CC SIGNAL 1 21 POTENTIAL.
CC FT CHAIN 22 151 MINOR CURLIN SUBUNIT.
CC SQ SEQUENCE 151 AA; 16182 MW; C0FC5430E6DD361D CRC64;

Query Match 13.7%; Score 106.5; DB 1; Length 151;
Best Local Similarity 30.7%; Pred. No. 0.015;
Matches 35; Conservative 17; Mismatches 43; Indels 19; Gaps 5;

QY 55 MAHALQSDARKSE-----TTITQSGYNGADVQ--GADNST-----IELTQNGFR 98
DB 18 IATATNYDLARSEYFVAVNELSKSFNCAAIIGVGTDSARVQRQSGKLLSVISQGEN 77
QY 99 NNATIDQWAKNSDIT-VGYGGNNALVNOTASDSSVMVRQVGFNNATANQY 151
DB 78 NRAKVQD--AGNYNFAYIEQTGNANDASISQAYGNSAALIQKSGNKANITQY 129

RESULT 7

ID YKO3 CABEL STANDARD; PRT; 401 AA.
AC P34291;
DT 01-FEB-1994 (Rel. 28; Created)
DT 30-MAY-2000 (Rel. 39; Last sequence update)
DT 30-MAY-2000 (Rel. 39; Last annotation update)
DE Hypothetical 43.5 kDa protein C05B5.3 in chromosome III.
GN C05B5.3
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Paloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Mortimore B.J.;
RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
RN [2]

RP REVISIONS.
RA Durbin R.;
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

CC EMBL; Z32679; CAA83596.1; -
CC PIR; C88571; C88571.
CC WormPep; C05B5.3; CE17369.
CC Hypothetical protein.
SQ SEQUENCE 401 AA; 43472 MW; C99DF259EF6C8B55 CRC64;

Query Match 11.8%; Score 92; DB 1; Length 401;
Best Local Similarity 23.2%; Pred. No. 0.6;
Matches 46; Conservative 19; Mismatches 71; Indels 62; Gaps 7;

```

QY 3 LLKVAFAALVWGSALAGVVPQWGG-----GNHNGGSSGPDYDQLVT 48
Db 30 LQYRAYAGSSGCVPAIVKSGFWPNADMIAGLQTEQSRQNSNNPQQDDPT 89
QY 49 RVVTHE-----MAHALQSDARKGETTI-----TQSG--YGNADVGGQA 85
Db 90 SQSTGQINGNVPGSSSSNQPVIIYIARAGSKYKNSVTTSTPTNGFNGFGQGNQ 149
QY 86 D-----NGTIELTQNGFRNATIDQWNAKNSDITVGOYGGNNAALVNOQTASD 132
Db 150 NTGFSGFFNNQNSQNLNNQNFQ-----QNLGASSGFNNQN-----QONSQQ 197
QY 133 SSVNVRQVGFNNATNQ 150
Db 198 NVNGPTSGFSNQTSNQ 215

RESULT 8
OMP63_RICJA STANDARD; PRT; 1656 AA.
AC O06653;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Outer membrane protein B precursor (168 kDa surface-layer protein)
DE (Surface protein antigen) (Cell surface antigen 5) (Sca5) (romp)
DE (romp B) (Contains: 120 kDa surface-exposed protein (Surface protein
DE antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide).
GN OMPB.
OS Rickettsia japonica.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=35790;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YH;
RA Uchiyama T.;
RT "Sequencing of the gene encoding the protein romp B of Rickettsia
RL japonica."
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THE 120 kDa SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (By
CC similarity).
CC -!- FUNCTION: THE 32 kDa BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR
CC (By similarity).
CC -!- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-
CC layer with hexagonal symmetry.
CC -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMP63 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB003681; BAA20138.1; -
CC InterPro; IPR006315; Autotransporter.
CC Pfam; PF03797; Autotransporter; 1.
CC TIGRFAMs; TIGR01414; autotrans_ba1; 2.
CC Antigen; S-layer; Cell wall.
CC CHAIN 1 1338 120 kDa SURFACE-EXPOSED PROTEIN.
CC FT CHAIN 1339 1656 32 kDa BETA PEPTIDE.
CC FT DOMAIN 528 533 POLY-GLY.
CC SQ SEQUENCE 1656 AA; 168097 MW; 3132A69C9DD5999F CRC64;
Query Match 11.8%; Score 92; DB 1; Length 1656;
Best Local Similarity 26.2%; Pred. No. 4;
Matches 45; Conservative 19; Mismatches 52; Indels 56; Gaps 9;

```

```

QY 6 VAAFAALVWGSALAGVVPQWGGGNGHNGGSSGPDYDQLVTRVVTHEMAHALQSDARK 65
Db 509 VLAGAILTQSATI-----TGDIHGSGGAA-----LQSI-----LANDATK 547
QY 66 SETTITQSG-----YNGADVQGGADNSTIELTQNGFRNATID----- 104
Db 548 ---TLTGGANIISANGGTINFQANGGTIKLTST--QNNIWDCLALAIADTQGVVDASS 602
QY 105 QWNAKNSDI--TVGOYGGNNAAL-----VNQTSDDSSVWVQVGFNNAT 147
Db 603 LTNAQTITISGTITIGANNNTLGGQFNGSSKTLGNGVAINELVIGNNGS 654

RESULT 9
GP63_LEIDO STANDARD; PRT; 590 AA.
AC P23223;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Leishmanolysin precursor (EC 3.4.24.36) (Cell surface protease)
DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface
DE endopeptidase).
GN GP63.
OS Leishmania donovani.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5661;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LV9;
RX MEDLINE=92107220; PubMed=1762629;
RA Webb J.R., Button L.L., McMaster R.W.;
RT "Heterogeneity of the genes encoding the major surface glycoprotein
RL of Leishmania donovani."
RL Mol. Biochem. Parasitol. 48:173-184(1991).
CC -!- FUNCTION: Has an integral role during the infection of macrophages
CC in the mammalian host.
CC -!- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and
CC P1' and basic residues at P2 and P3'. A model nonapeptide is
CC cleaved at -Ala-Tyr--Leu-Lys-Lys-.
CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -!- SIMILARITY: Belongs to peptidase family M8.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M60048; AAA29244.1; -
CC HSP; P08148; 1LML.
CC DR MEROPS; M08.001; -
CC DR InterPro; IPR006025; Pept M Zn Bs.
CC DR InterPro; IPR001577; Peptidase_M8.
CC DR Pfam; PF01457; Peptidase_M8; 1.
CC DR PRINTS; PR00782; LSHMANOLYSIN.
CC DR PROSITE; PS00142; ZINC_PROTEASE; 1.
CC KW Hydrolyase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;
CC Zymogen; Signal; Cell adhesion; GPI-anchor; Lipoprotein.
CC FT SIGNAL 1 39 POTENTIAL.
CC FT PROPEP 40 87 ACTIVATION PEPTIDE.
CC FT CHAIN 88 565 LEISHMANOLYSIN.
CC FT PROPEP 566 590 REMOVED IN MATURE FORM (BY SIMILARITY).
CC FT METAL 251 251 ZINC (CATALYTIC) (BY SIMILARITY).
CC FT ACT_SITE 252 252 BY SIMILARITY.
CC FT METAL 255 255 ZINC (CATALYTIC) (BY SIMILARITY).
CC FT METAL 321 321 ZINC (CATALYTIC) (BY SIMILARITY).
CC FT DISULFID 112 129 BY SIMILARITY.
CC FT DISULFID 178 217 BY SIMILARITY.
CC FT DISULFID 301 373 BY SIMILARITY.

```


-!- SUBCELLULAR LOCATION: Nuclear at specific periods of the day.
First accumulates in the perinuclear region about one hour before translocation into the nucleus. Interaction with Tim is required for nuclear localization (By similarity).

-!- PTM: PHOSPHORYLATED WITH A CIRCADIAN RHYTHMICITY. PROBABLY BY THE DOUBLE-TIME PROTEIN (DBT). PHOSPHORYLATION COULD BE IMPLICATED IN THE STABILITY OF PER MONOMER AND IN THE FORMATION OF HETERODIMER PER-TIM (BY SIMILARITY).

-!- MISCELLANEOUS: THE SEQUENCE SHOWN IS THAT OF STRAIN ATUXXO.

-!- SIMILARITY: Contains 2 PAS (PER-ARNT-SIM) dimerization domains.

-!- SIMILARITY: Contains 1 PAS-associated C-terminal (PAC) domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).

FT VARIANT 734 734 MISSING (IN STRAINS MANAUS 4, PORTO
 FT VARIANT 747 747 ALEGRE 1 AND PORTO ALEGRE 2).
 FT VARIANT 764 766 S -> A (IN STRAINS GUADELOUPE AND GUANA).
 FT VARIANT 886 886 MISSING (IN STRAIN MANAUS 3).
 FT NON-TER 1093 1093 A -> T (IN STRAIN 0811.4).
 SQ SEQUENCE 1093 AA; 115996 MW; AB6DE050267EC187 CRC64;
 Query Match 11.7%; Score 91; DB 1; Length 1093;
 Best Local Similarity 26.1%; Pred. No. 3.1;
 Matches 24; Conservative 8; Mismatches 14; Gaps 2;
 QY 27 GGGGNGGNGSGPPDYDQLVTRVTHMAHALQSDARKSETITQSGVNGADVGGQAD 86
 DB 719 GGGGGGGGGGGGGGGGLPLFL-----DVTHTSSSSQNKGPCTGVAAGGAGGGVGGGG-- 770
 QY 87 NSTIETONGFRNATIDQWNAKNNSDITVGOY 118
 DB 771 -----SCSLGNGNGVSGNGNSQPSINQY 796
 RESULT 11
 GP63 LEIMA
 ID GP63 LEIMA STANDARD; PRT; 602 AA.
 AC P08148; P15906;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Leishmanolysin precursor (EC 3.4.24.36) (Cell surface protease)
 DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface
 DE endopeptidase).
 GN GP63
 OS Leishmania major.
 OC Eukaryota; Eulenzozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 OX NCBI_TaxID=5664;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 101-123.
 RX MEDLINE=88154764; PubMed=3346625;
 RA Button L.L., McMaster W.R.;
 RT "Molecular cloning of the major surface antigen of leishmania.";
 RL J. Exp. Med. 167:724-729(1988).
 RN [2]
 RP REVISIONS.
 RA Button L.L., McMaster W.R.;
 RL J. Exp. Med. 171:589-589(1990).
 RN [3]
 RP GPI-ANCHOR.
 RX MEDLINE=91009116; PubMed=2145267;
 RA Schneider P., Ferguson M.A.J., McConville M.J., Mehler A.,
 RA Homans S.W., Bordier C.;
 RT "Structure of the glycosyl-phosphatidylinositol membrane anchor of
 RT the Leishmania major promastigote surface protease.";
 RL J. Biol. Chem. 265:16955-16964(1990).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).
 RX MEDLINE=95406217; PubMed=7675788;
 RA Schlagenhauf E., Etges R., Metcalf P.;
 RT "Crystallization and preliminary X-ray diffraction studies of
 RT leishmanolysin, the major surface metalloproteinase from Leishmania
 RT major.";
 RL Proteins 22:58-66(1995).
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY (1.86 ANGSTROMS).
 RX MEDLINE=98416698; PubMed=9739094;
 RA Schlagenhauf E., Etges R., Metcalf P.;
 RT "The crystal structure of the Leishmania major surface proteinase
 RT leishmanolysin.";
 RL Structure 6:1035-1046(1998).
 CC -!- FUNCTION: Has an integral role during the infection of macrophages
 CC in the mammalian host.
 CC -!- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and
 CC P1' and basic residues at P2 and P3'. A model nonapeptide is
 CC cleaved at -Ala-Tyr-|-Leu-Lys-Lys-.

CC -!- COPACITOR: Binds 1 zinc ion per subunit.
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
 CC -!- PM: THE PHOSPHATIDYLINOSITOL MOIETY OF THE GPI-ANCHOR CONTAINS A
 CC FULLY SATURATED, UNBRANCHED 1-O-ALKYL CHAIN (MAINLY C24:0) AND A
 CC MIXTURE OF FULLY SATURATED UNBRANCHED 2-O-ACYL CHAINS (C12:0,
 CC C14:0, C16:0, AND C18:0).
 CC -!- SIMILARITY: Belongs to peptidase family M8.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@isb-sib.ch).
 CC ENBL; Y00647; CAA68673.1; -;
 DR PIR; P02221; P02221.
 DR PDB; 1LMD; 17-SEP-97.
 DR MEROPS; M08.001; -;
 DR InterPro; IPR006025; Pept_M_Zn_BS.
 DR InterPro; IPR001577; Peptidase_M8.
 DR Pfam; PF01457; Peptidase_M8; 1.
 DR PRINTS; PR00782; LSHMANOLYSIN.
 DR PROSITE; PS00142; ZINC_PROTEASE; 1.
 DR KW Hydrolase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;
 FT SIGNAL 1 39 POTENTIAL.
 FT PROPEP 40 100 ACTIVATION PEPTIDE.
 FT CHAIN 101 577 LEISHMANOLYSIN.
 FT PROPEP 578 602 REMOVED IN MATURE FORM.
 FT METAL 264 264 ZINC (CATALYTIC).
 FT ACT_SITE 265 265 ZINC (CATALYTIC).
 FT METAL 268 268 ZINC (CATALYTIC).
 FT METAL 334 334 ZINC (CATALYTIC).
 FT DISULFID 125 142
 FT DISULFID 191 230
 FT DISULFID 314 386
 FT DISULFID 393 455
 FT DISULFID 406 425
 FT DISULFID 415 489
 FT DISULFID 466 510
 FT DISULFID 515 565
 FT DISULFID 535 558
 FT CARBOHYD 300 300
 FT CARBOHYD 407 407
 FT LIPID 577 577
 FT STRAND 101 102
 FT STRAND 107 108
 FT STRAND 111 114
 FT HELIX 116 119
 FT TURN 121 122
 FT TURN 128 129
 FT STRAND 131 133
 FT STRAND 139 141
 FT HELIX 144 146
 FT HELIX 150 158
 FT TURN 159 159
 FT HELIX 160 169
 FT TURN 170 171
 FT STRAND 172 174
 FT STRAND 177 178
 FT STRAND 180 181
 FT TURN 189 190
 FT HELIX 191 193
 FT HELIX 198 202
 FT TURN 203 203
 FT STRAND 205 206
 FT STRAND 210 215
 FT TURN 221 222
 FT STRAND 226 232
 FT TURN 234 235
 FT STRAND 238 244
 N-LINKED (GLCNAC. .) (POTENTIAL).
 N-LINKED (GLCNAC. .) (POTENTIAL).
 GPI-anchor amidated asparagine.

DR PROSITE; PS00142; ZINC_PROTEASE.1
KW Hydrolase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;
KW Zymogen; Signal; Cell adhesion; GPI-anchor; Lipoprotein.
FT SIGNAL 1 39 POTENTIAL.
FT PROPEP 40 97 ACTIVATION PEPTIDE.
FT CHAIN 98 574 LEISHMANOLYSIN.
FT FT 575 599 REMOVED IN NATURE FORM (BY SIMILARITY).
FT METAL 261 261 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 262 262 BY SIMILARITY.
FT METAL 265 265 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 331 331 ZINC (CATALYTIC) (BY SIMILARITY).
FT DISULFID 122 139 BY SIMILARITY.
FT DISULFID 311 383 BY SIMILARITY.
FT DISULFID 390 452 BY SIMILARITY.
FT DISULFID 403 422 BY SIMILARITY.
FT DISULFID 412 486 BY SIMILARITY.
FT DISULFID 463 507 BY SIMILARITY.
FT DISULFID 512 562 BY SIMILARITY.
FT DISULFID 532 555 BY SIMILARITY.
FT CARBOHYD 297 297 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 394 394 N-LINKED (GLCNAC. .) (POTENTIAL).
FT LIPID 574 574 GPI-anchor amidated asparagine (By similarity)
SQ SEQUENCE 599 AA; 63848 MW; 746730AE8E2A2E7C CRC64;

Query Match 11.3%; Score 88; DB 1; Length 599;
Best Local Similarity 46.7%; Pred. No. 2.8;
Matches 21; Conservative 3; Mismatches 11; Indels 10; Gaps 1;

QY 43 YDLVTRVVTHMAHALQSDARKSETITQSGYGNGADVCGGADN 87
||||| : : : : |
Db 251 YDLVTRVVTHMAHALGFVS-----GFEGARILEISGN 285

RESULT 15
OVO_DROME STANDARD; PRT; 1028 AA.

ID OVO_DROME STANDARD; PRT; 1028 AA.
AC P51521; Q9XZU4;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ovo protein (Shaven baby protein).
GN OVO OR SVB.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
CC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
CC Ephydroidea; Drosophilidae; Drosophila.
CX NCBI_TaxID=7227;
RN [1]
RS SEQUENCE FROM N.A.
RP TISSUE=Ovary;
RC MEDLINE=95051209; PubMed=7935398;
RX Garfinkel M.D., Wang J., Liang Y., Mahowald A.P.;
RA "Multiple products from the shavenbaby-ovo gene region of Drosophila melanogaster: relationship to genetic complexity."; Mol. Cell. Biol. 14:6809-6818(1994). [2]

RN RN
RS SEQUENCE FROM N.A.
RP STRAIN=Oregon-R;
RX MEDLINE=91293102; PubMed=1712294;
RA Mevel-Ninio M.T.M., Terracol R., Kafatos F.C.;
RT "The ovo gene of Drosophila encodes a zinc finger protein required for female germ line development."; EMBO J. 10:2259-2266(1991).

CC CC -1- FUNCTION: REQUIRED FOR SURVIVAL AND DIFFERENTIATION OF FEMALE GERM LINE CELLS. PLAYS A ROLE IN GERM LINE SEX DETERMINATION.
CC CC -1- SUBCELLULAR LOCATION: Nuclear (potential).
CC CC -1- DEVELOPMENTAL STAGE: FIRST APPEARS IN THE GERMARIIUM AND ACCUMULATES IN NURSE CELLS DURING OOGENESIS. STORED IN THE EGG, BUT IS RAPIDLY LOST IN THE EMBRYOS EXCEPT FOR ITS CONTINUED PRESENCE IN THE GERM LINE PRECURSOR POLE CELLS.
CC CC -1- SIMILARITY: Contains 4 C2H2-type zinc fingers.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; U11383; AAB50216.1; -.
EMBL; X59772; CAB36921.1; ALT_SEQ.
PIR; A56038; A56038.
HSP; P07248; 2ADR.
TRANSFAC; T00669; -.
FlyBase; FBgn003028; ovo.
InterPro; IPR007087; znf_C2H2.
Pfam; PF00936; zf-C2H2; 3.
SMART; SM00355; znf_C2H2; 4.
PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
PROSITE; PS0157; ZINC_FINGER_C2H2_2; 3.
KW Zinc-finger; Metal-binding; Repeat; Nuclear protein;
KW Transcription regulation.
DOMAIN 62 66 POLY-ALA.
FT DOMAIN 72 77 POLY-GLY.
FT DOMAIN 80 85 POLY-GLY.
FT DOMAIN 98 108 POLY-GLY.
FT DOMAIN 144 152 POLY-HIS.
FT DOMAIN 153 159 POLY-ASN.
FT DOMAIN 336 339 POLY-GLN.
FT DOMAIN 347 353 POLY-GLN.
FT DOMAIN 357 361 POLY-GLN.
FT DOMAIN 410 414 POLY-GLN.
FT DOMAIN 418 422 POLY-GLN.
FT DOMAIN 426 432 POLY-GLN.
FT DOMAIN 445 453 POLY-GLN.
FT DOMAIN 456 459 POLY-GLN.
FT DOMAIN 466 474 POLY-GLN.
FT DOMAIN 497 517 POLY-ALA.
FT DOMAIN 524 529 POLY-SER.
FT DOMAIN 549 558 POLY-ALA.
FT DOMAIN 639 651 POLY-ALA.
FT DOMAIN 717 725 POLY-ALA.
FT DOMAIN 797 802 POLY-GLN.
FT DOMAIN 820 823 POLY-GLN.
FT DOMAIN 826 832 POLY-GLN.
FT 2N FING 874 896 C2H2-TYPE 1.
FT ZN_FING 902 924 C2H2-TYPE 2.
FT ZN_FING 930 953 C2H2-TYPE 3.
FT ZN_FING 969 992 C2H2-TYPE 4.
FT CONFLICT 647 647 A -> R (in REF. 2).
SQ SEQUENCE 1038 AA; 110620 MW; D7068BB2BC0F6F77 CRC64;

Search completed: August 2, 2004, 14:49:29
Job time : 5.3 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:39:17 ; Search time 29.7 Seconds
(without alignments)
1604.150 Million cell updates/sec

Title: US-09-543-407-18
Perfect score: 780
Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVFGNNATANAQY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL 25: *
1: sp_archaea: *
2: sp_bacteria: *
3: sp_fungi: *
4: sp_human: *
5: sp_invertebrate: *
6: sp_mammal: *
7: sp_mhc: *
8: sp_organelle: *
9: sp_phage: *
10: sp_plant: *
11: sp_rodent: *
12: sp_virus: *
13: sp Vertebrate: *
14: sp Unclassified: *
15: sp_rvirus: *
16: sp_bacteriap: *
17: sp_archaeap: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	682	87.4	152	2	Q33802 salmoneilla
2	593.5	76.1	150	2	Q7x243 citrobacter
3	550	70.5	149	2	Q7x240 citrobacter
4	498.5	63.9	152	16	Q8cW63 escherichia
5	428.5	54.9	150	2	Q7x237 enterobacte
6	303	38.8	76	2	Q54069 salmoneilla
7	122	15.6	29	2	Q8S3U5 escherichia
8	114	14.6	1748	5	Q94821 tetrahymena
9	113	14.5	502	16	Q8eiH4 shewanella
10	110.5	14.2	151	2	Q7x244 citrobacter
11	108	13.8	171	16	Q89J13 bradyrhizob
12	107.5	13.8	151	16	Q7UCZ1 shigella fl
13	107.5	13.8	160	16	Q8cW64 escherichia
14	107.5	13.8	160	16	Q83RU7 shigella fl
15	106.5	13.7	1209	16	Q89CK5 bradyrhizob
16	105.5	13.5	91	2	Q9S3J8 escherichia

17	101.5	13.0	154	16	Q89J15	Q89J15 bradyrhizob
18	101	12.9	262	5	Q9VIX5	Q9VIX5 drosophila
19	100	12.8	179	2	Q33801	Q33801 salmoneilla
20	99	12.7	130	16	Q89J14	Q89J14 bradyrhizob
21	98	12.6	157	16	Q88HG0	Q88HG0 pseudomonas
22	98	12.6	362	16	Q8EV84	Q8EV84 mycoplasma
23	98	12.6	3659	16	Q98LN6	Q98LN6 rhizobium 1
24	97.5	12.5	151	2	Q7x238	Q7x238 enterobacte
25	97.5	12.5	152	2	Q7x241	Q7x241 citrobacter
26	97	12.4	368	16	Q8EWD6	Q8EWD6 mycoplasma
27	97	12.4	1422	16	Q8EFU3	Q8EFU3 shewanella
28	97	12.4	2035	2	Q9XCJ4	Q9XCJ4 salmoneilla
29	97	12.4	2039	16	Q8ZN57	Q8ZN57 salmoneilla
30	96.5	12.4	1765	16	Q7V855	Q7V855 prochloroco
31	96	12.3	490	16	Q8EY19	Q8EY19 leptospira
32	95.5	12.2	145	16	Q8U6N9	Q8U6N9 agrobacteri
33	95.5	12.2	1613	2	Q9KKB2	Q9KKB2 israeli tic
34	95.5	12.2	3501	16	Q8Y106	Q8Y106 ralstonia s
35	95.5	12.2	3552	16	Q8XSD6	Q8XSD6 ralstonia s
36	94.5	12.1	153	16	Q89J16	Q89J16 bradyrhizob
37	94.5	12.1	348	13	Q93397	Q93397 cyprinus ca
38	94	12.1	480	16	Q89EV2	Q89EV2 bradyrhizob
39	93.5	12.0	139	16	Q8EIH3	Q8EIH3 shewanella
40	93.5	12.0	287	5	Q9VIX6	Q9VIX6 drosophila
41	93.5	12.0	582	16	P71868	P71868 mycobacteri
42	93.5	12.0	582	16	Q7TW98	Q7TW98 mycobacteri
43	93.5	12.0	1615	2	Q9KKA8	Q9KKA8 rickettsia
44	93	11.9	141	16	Q8U6P1	Q8U6P1 agrobacteri
45	92	11.8	353	16	Q8EV92	Q8EV92 mycoplasma

ALIGNMENTS

RESULT 1

O33802 ID C33802 PRELIMINARY; PRT; 152 AA.
AC O33802;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE AgfA protein (Fragment).
GN AGFA.
OS Salmoneilla typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmoneilla.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98053981; PubMed=9393832;
RA Sukupolvi S.S., Lorentz R.G., Gordon J.I., Bian Z., Pfeifer J.D.,
RA Normark S.J., Rhen M.;
RT "Expression of thin, aggregative fimbriae promotes interaction of
RT Salmoneilla typhimurium SR-11 with mouse small intestinal epithelial
RT cells."
RL Infect. Immun. 65:5320-5325 (1997).
DR EMBL; AJ000514; CAA04151.1; -.
FT NON TER 152 152
SQ SEQUENCE 152 AA; 15401 MW; 9DA7DADC2364B006 CRC64;

Query Match 87.4%; Score 682; DB 2; Length 152;
Best Local Similarity 89.4%; Pred. No. 1.7e-46;
Matches 135; Conservative 4; Mismatches 12; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSSGPDYDOLVTRVVTHEMAHALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSSGPDSTLSIYGYGSANAALAQ 60

QY 61 SPARKSETTITSGYNGADVCGGADNSTIETQTGFRNNATIDOWNAKNSITTYGVGG 120
Db 61 SPARKSETTITSGYNGADVCGGADNSTIETQTGFRNNATIDOWNAKNSITTYGVGG 120
QY 121 NNAALVNTQASDSSVMVRQVGFNNATANAQY 151

```

Db      121 NNAALVNQTASDSSVMVROVGFNNAPANQY 151
|||||
RESULT 2
Q7X243
ID Q7X243 PRELIMINARY; PRT; 150 AA.
AC Q7X243;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DE Curlin-csgA protein.
GN CSGA.
OS Citrobacter sp. Fec2.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Citrobacter.
OX NCBI_TaxID=213763;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fec2;
RA Zogaj X., Bokranz W., Nimtz M., Romling U.;
RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RT Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158 (2003).
DR EMBL; AJ515700; CAD56672.1; -.
SQ SEQUENCE 150 AA; 15016 MW; 1D7141B8D6973DC6 CRC64;

Query Match 76.1%; Score 593.5; DB 2; Length 150;
Best Local Similarity 78.1%; Pred. No. 1.7e-39;
Matches 118; Conservative 14; Mismatches 18; Indels 1; Gaps 1;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGGPDYDQLVTRVVTHEMAHALQ 60
|||||
Db 1 MKLLQVAAFAAIVVSGSALAGVVPQWGGGG-GGGSSSGPESTLSIYQSGVNAALALQ 59
|||||
QY 61 SDARKSETTITQSGYNGADVCGGADNSITELTQNGFNNATIDQWNAKNSDITVQYGG 120
|||||
Db 60 SDARKSDTTIHQNGFGNGADVCGGSDNSTIDLTQNGFKNATIDQWNGKNSDITVQYGG 119
|||||
QY 121 NNAALVNQTASDSSVMVROVGFNNATANQY 151
|||||
Db 120 HNAALVNQTASDSSVLVHVQVGFNNATANQY 150
|||||

RESULT 3
Q7X240
ID Q7X240 PRELIMINARY; PRT; 149 AA.
AC Q7X240;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DE Curlin-csgA protein.
GN CSGA.
OS Citrobacter freundii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Citrobacter.
OX NCBI_TaxID=546;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fec4;
RA Zogaj X., Bokranz W., Nimtz M., Romling U.;
RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RT Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158 (2003).
DR EMBL; AJ515701; CAD56675.1; -.
SQ SEQUENCE 149 AA; 15260 MW; 946DD52017F648FD CRC64;

Query Match 70.5%; Score 550; DB 2; Length 149;
Best Local Similarity 72.2%; Pred. No. 4.5e-36;
Matches 109; Conservative 20; Mismatches 20; Indels 2; Gaps 1;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGGPDYDQLVTRVVTHEMAHALQ 60
|||||

```

```

Db      1 MKLLKVAAPAAIVVSGSALAGVVPQW--GGNHGGSNGYGPDSLSIYQYGSNNSANALQ 58
QY 61 SDARKSETTITQSGYNGADVCGGADNSITELTQNGFNNATIDQWNAKNSDITVQYGG 120
|||||
Db 59 SDARKSDVTITQHRGNGAVVQGGADDSISLKTQGFQNSATIDQWNAKNSDITVQFGG 118
|||||
QY 121 NNAALVNQTASDSSVMVROVGFNNATANQY 151
|||||
Db 119 RNCALVNQTASDSSVLIQVGFNNATANQH 149
|||||

RESULT 4
Q8CW63
ID Q8CW63 PRELIMINARY; PRT; 152 AA.
AC Q8CW63;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Major curlin subunit precursor.
GN CSGA OR C1306.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).
DR EMBL; AC016759; AAN79779.1; -.
KW Complete proteome.
SQ SEQUENCE 152 AA; 15064 MW; 3BA57F34C1240E83 CRC64;

Query Match 63.9%; Score 498.5; DB 16; Length 152;
Best Local Similarity 67.1%; Pred. No. 5.4e-32;
Matches 102; Conservative 20; Mismatches 29; Indels 1; Gaps 1;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQW--GGGNGHNGGNSGGPDYDQLVTRVVTHEMAHAL 59
|||||
Db 1 MKLLKVAAPAAIVVSGSALAGVVPQYGGGNGHNGGNSGNSSELNIYQYGGNSALAQ 60
|||||
QY 60 QSDARKSETTITQSGYNGADVCGGADNSITELTQNGFNNATIDQWNAKNSDITVQYGG 119
|||||
Db 61 QADARNSDLTITQHGNGADVCGGSDSSILDTQNGFNSATLDQWNGKNSDITVQKQFG 120
|||||
QY 120 GNNALVNQTASDSSVMVROVGFNNATANQY 151
|||||
Db 121 GNGRAVDQTASNSSVNVTVQVGFNNATAHQY 152
|||||

RESULT 5
Q7X237
ID Q7X237 PRELIMINARY; PRT; 150 AA.
AC Q7X237;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Curlin-csgA protein.
GN CSGA.
OS Enterobacter sakazakii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Enterobacter.
OX NCBI_TaxID=28141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fec39;
RA Zogaj X., Bokranz W., Nimtz M., Romling U.;

```



```

OX NCBI_TaxID=562;
RN [1]
RN SEQUENCE FROM N.A.
RP TRANSPOS=Insertion sequence IS1;
RC MEDLINE=99314153; PubMed=10386375;
RX La Ragione R.M., Collighan R.J., Woodward M.J.;
RT "Non-curliation of Escherichia coli O78:K80 isolates associated with
RT IS1 inserti on in cs9B and reduced persistence in poultry infection.";
RL FEMS Microbiol. Lett. 175:247-253(1999).
RD EMBL; AJ131756; CAB45380.1; -
RE NON TER 29
SF SEQUENCE 29 AA; 2789 MW; E290DFC07ABBE243 CRC64;

Query Match 15.6%; Score 122; DB 2; Length 29;
Best Local Similarity 89.7%; Pred.No.0.0042; 2; Indels 0; Gaps 0;
Matches 26; Conservative 1; Mismatches 2;

QY 1 MKLLKVAFAAIIVYVSGSALAGVVPQMGGG 29
DB 1 MKLLKVAATAAIVFSGSALAGVVPQYGGG 29

RESULT 8
Q94821 PRELIMINARY; FRT; 1748 AA.
ID Q94821; P92146; P92145; P92144; P92143; P92142; P92141; Q94820;
AC Q94821; P92146; P92145; P92144; P92143; P92142; P92141; Q94820;
DT 01-FEB-1997 (T-EMBLrel. 02, Created)
DT 01-FEB-1997 (T-EMBLrel. 02, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE CNJB protein.
GN CNJB.
OS Tetr hymena thermophila.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
OC Tetr hymenina; Tetr hymena.
OX NCBI_TaxID=5911;
RN [1]
RN SEQUENCE FROM N.A.
RP MEDLINE=88189811; PubMed=3357771;
RX Martindale D.W., Taylor F.M.;
RA "Multiple introns in a conjugation-specific gene from Tetr hymena
RT thermophila.";
RL Nucleic Acids Res. 16:2189-2201(1988).
RE NON TER 29
SF SEQUENCE FROM N.A.
RP MEDLINE=94051569; PubMed=8233799;
RX Taylor F.M., Martindale D.W.;
RT "Retroviral-type zinc fingers and glycine-rich repeats in a protein
RT encoded by cnbB, a Tetr hymena gene active during meiosis.";
RL Nucleic Acids Res 21:4610-4614(1993).
RD EMBL; X06462; CAB37323.1; -
DR EMBL; LC3710; AAC37171.1; -
DR FIR; S42136; S42136.
DR HSSP; P05888; 1AAP.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001878; Znf.CCHC.
DR Pfam; PF00098; zf.CCHC; 7.
DR PRINTS; PR00939; C2HCZNFINGER.
DR SMART; SM00343; Znf.C2HC; 7.
DR PROSITE; PS00158; ZF_CCHC; 7.
FT CONFLICT 251 251 M -> I (IN REF. 1).
FT CONFLICT 256 256 I -> N (IN REF. 1).
SF SEQUENCE 1748 AA; 199524 MW; 0B03F210104008A3 CRC64;

Query Match 14.6%; Score 114; DB 5; Length 1748;
Best Local Similarity 30.5%; Pred.No.2.3;
Matches 39; Conservative 22; Mismatches 27; Indels 40; Gaps 8;

QY 25 QXGGGNNHNGG--GNSSGPDYDQLVTRVVTHEMAHALOSDARKSTTTTQSGVGNQADV 81
DB 1640 QFGGGNSNGGSGWGTSSTSSDWN-----CQSNVOES-ITTSGGWGS----- 1680

QY 82 GCGADNSTIELTQNGFRNNATIDQWNAKNSDITVGYQGNNAALVYNQATSDSSVWVRQVG 141

```

```

Db 1681 -SGSGNQ-----TGGWGSN---DNQQQQQVNTGGGGSSNS---NQTNSS----- 1722
Qy 142 FGNNTAN 149
Db 1723 WGSNNQAS 1730

RESULT 9
Q8EIH4 PRELIMINARY; PRT; 502 AA.
ID Q8EIH4
AC Q8EIH4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Conserved hypothetical protein.
GN SC0865.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.
OX NCBI_TaxID=70863;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MR-1;
RX MEDLINE=2297686; PubMed=1236813;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
RA Meyer T., Tsipin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
RA Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,
RA Mueller J., Khouri H., Gill J., Uterback T.R., McDonald L.A., C.M.;
RA Feldblyum T.V., Smith H.O., Venter J.C., Nelson K.H., Fraser C.M.;
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis.";
RL Nat. Biotechnol. 20:1118-1123(2002).
DR EMBL; AF015532; AAN53941.1; -.
DR TIGR; SC0865; -.
SQ Hypothetical protein; Complete proteome.

Query Match 14.5%; Score 113; DB 16; Length 502;
Best Local Similarity 26.0%; Pred. No. 0.64;
Matches 40; Conservative 20; Mismatches 58; Indels 36; Gaps 6;

Qy 29 GGNHNG-----GGN-----SSGPDYDQVTR-----VVTHEMAHAL 59
Db 231 GDNHTGFVYALAGSNDISMEEGSGNNNTAYLSMTTGDNTVDITODGDSNTVGSLSIADI 290
Qy 60 QSDARKSETTITQSGYNGADYVGQADNRTIELTQNGFRNNTATIDOWNAKNSDITVGYQY 119
Db 291 QGD--DNDITIKQKDSNGAEQVWGVSDNDVDLKGQGDANFATFGAYGTDN-DFDLSK 347
Qy 120 GNNALVNQTASDSSVMVRQVGFN-----NATAN 149
Db 348 DNNELVAFATGDNSEIETISQEGDANFAYVDATGN 381

RESULT 10
Q7X244 PRELIMINARY; PRT; 151 AA.
ID Q7X244
AC Q7X244;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Nucleation component of curlin monomers.
GN CSGB.
OS Citrobacter sp. Fec2.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Citrobacter.
OX NCBI_TaxID=213763;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fec2;

```

```

RA Zogaj X., Bokranz W., Nintz M., Rowling U.;
RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RT Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158(2003).
DR EMBL; AJ515700; CAD56671.1; -.
SQ SEQUENCE 151 AA; 16158 MW; BD00AF57E1400704 CRC64;

Query Match 14.2%; Score 110.5; DB 2; Length 151;
Best Local Similarity 25.0%; Pred. No. 0.24;
Matches 29; Conservative 26; Mismatches 46; Indels 15; Gaps 3;

Qy 47 VTRVVTHEMAHALQSDARKSETTITQSGYNGADYVGQADNRTIELTQNGFR----- 98
Db 18 IASATSYDLAH---SEYNFAVNELSKSFNQAAITIGVGTNNSAKMRQSGSKLLSVSGE 74
Qy 99 ---NNATIDOWNAKNSDITVQYGGNNAALVNQTASDSSVMVRQVGFNNTATANQY 151
Db 75 GGSNRAKVDSGAYNF-AVIAQSGHSDASISQSNVYGTAMIIQKSGNKANITQY 129

RESULT 11
Q89JI3 PRELIMINARY; PRT; 171 AA.
ID Q89JI3
AC Q89JI3;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE CSGA protein.
GN CSGA OR BLL5300.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RX MEDLINE=22484998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA Sasamoto S., Watanabe A., Idesawa K., Iriguchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimo S., Tsurucka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110.";
RL DNA Res. 9:189-197(2002).
DR EMBL; AP005954; BAC50565.1; -.
KW Complete proteome.
SQ SEQUENCE 171 AA; 17448 MW; 99SDB08C01498381 CRC64;

Query Match 13.8%; Score 108; DB 16; Length 171;
Best Local Similarity 36.6%; Pred. No. 0.44;
Matches 34; Conservative 10; Mismatches 39; Indels 10; Gaps 3;

Qy 58 ALQSDARKSETTITQSGYNGADYVGQADNRTIELTQNGFRNNTATIDOWNAKNSDITV 115
Db 52 ALSSAAQAANTSTTVQVGLVNGSVTQNGLTNDSSSTTQIGILNGASTWGTSSPS---- 107
Qy 116 GOYGGNNAALVNQTASDSSVMVRQVGFNNTA 148
Db 108 ---LNNVSTVQAGVQNSATTGQVAFGNGSA 136

RESULT 12
Q7UCZ1 PRELIMINARY; PRT; 151 AA.
ID Q7UCZ1
AC Q7UCZ1;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Minor curlin subunit.
GN CSGB OR S1108.
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.

```

```

ON NCBI_TaxID=623;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN=2457T / ATCC 700930 / Serotype 2a;
RX MEDLINE=2259274; PubMed=12704152;
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
RA Schwartz D.C., Blattner F.R.;
RT "Complete genome sequence and comparative genomics of Shigella
RT flexneri serotype 2a strain 2457T.";
RL Infect. Immun. 71:2775-2786(2003).
DR EMBL; AE016981; AAP16542.1; -.
SQ SEQUENCE 151 AA; 15868 MW; 5D5D266B964014A0 CRC64;

Query Match 13.8%; Score 107.5; DB 16; Length 151;
Best Local Similarity 35.4%; Pred. No. 0.45;
Matches 29; Conservative 9; Mismatches 43; Indels 1; Gaps 1;

QY 70 ITQSGYNGADVGGADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGGNNAALVNQT 129
Db 49 IGQAGTNSAQLRGGSKLLAVVAQEGSSNRKIDQTGYNL-AYIDQAGSANDASISQ 107

QY 130 ASDSSVMVRQVGFNNATANY 151
Db 108 AYGNTAMIIQKSGNKANITQY 129

RESULT 13
Q8CW64
ID Q8CW64 PRELIMINARY; PRT; 160 AA.
AC Q8CW64;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Minor curlin subunit precursor.
GN CSGB OR C1303.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D.A., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
DR EMBL; AE016759; AAN79778.1; -.
KW Complete proteome.
SQ SEQUENCE 160 AA; 16963 MW; 49F68448D979B986 CRC64;

Query Match 13.8%; Score 107.5; DB 16; Length 160;
Best Local Similarity 35.4%; Pred. No. 0.45;
Matches 29; Conservative 9; Mismatches 43; Indels 1; Gaps 1;

QY 70 ITQSGYNGADVGGADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGGNNAALVNQT 129
Db 58 IGQAGTNSAQLRGGSKLLAVVAQEGSSNRKIDQTGYNL-AYIDQAGSANDASISQ 116

QY 130 ASDSSVMVRQVGFNNATANY 151
Db 117 AYGNTAMIIQKSGNKANITQY 138

RESULT 14
Q83RU7
ID Q83RU7 PRELIMINARY; PRT; 160 AA.
AC Q83RU7;
```

```

DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Minor curlin subunit precursor, similar to csgA.
GN CSGB OR SF1035.
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN=301 / Serotype 2a;
RX MEDLINE=22272406; PubMed=12384590;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Yu J.;
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RT through comparison with genomes of Escherichia coli K12 and O157.";
RL Nucleic Acids Res. 30:4432-4441(2002).
DR EMBL; AB015131; AAN42658.1; -.
KW Complete proteome.
SQ SEQUENCE 160 AA; 16919 MW; 50269F5268D2A32F CRC64;

Query Match 13.8%; Score 107.5; DB 16; Length 160;
Best Local Similarity 35.4%; Pred. No. 0.45;
Matches 29; Conservative 9; Mismatches 43; Indels 1; Gaps 1;

QY 70 ITQSGYNGADVGGADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGGNNAALVNQT 129
Db 58 IGQAGTNSAQLRGGSKLLAVVAQEGSSNRKIDQTGYNL-AYIDQAGSANDASISQ 116

QY 130 ASDSSVMVRQVGFNNATANY 151
Db 117 AYGNTAMIIQKSGNKANITQY 138

RESULT 15
Q89CK5
ID Q89CK5 PRELIMINARY; PRT; 1209 AA.
AC Q89CK5;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE BIL7792 protein.
GN BIL7792;
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RX MEDLINE=22484998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA Sasamoto S., Watanabe A., Idesawa K., Iriguchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimpo S., Tsurucka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110.";
RL DNA Res. 9:189-197(2002).
DR EMBL; AP005963; BAC53057.1; -.
KW Complete proteome.
SQ SEQUENCE 1209 AA; 118292 MW; 50F97581D524EB71 CRC64;

Query Match 13.7%; Score 106.5; DB 16; Length 1209;
Best Local Similarity 25.0%; Pred. No. 5.9;
Matches 47; Conservative 20; Mismatches 66; Indels 55; Gaps 8;

QY 9 FPAIVVSGSALAGVVVQWGGGNGHNGGNSCPDYLQLVTRVVTHEMAH----ALQSDAR 64
Db 934 YGNVTVTGSTGATTIAL-----GNGNDTIDASG--YGNVITLGNNGNDIVHPGDGASQTAG 987
```

QY	65	KSETTTQSGYGNDVGGQAD-----NSTELTQGNFRNATIDQWN-	107
DB	988	NGNDLVTLGSGYNTVLMGNDVVAGDGAANSVTLGDGNNVTNL--GGMCNQITVSGGNTN	1045
QY	108	----AKNSDITVQYQ-----GNNAAALVNQ-----TASDSSVMRVQVQ	141
DB	1046	AIIAGSGSNVAVAGAGHDTIMLGGAANHVVVLGSGQANVTNIQIGDQVVTVYNGGSDQFNFG	1105
QY	142	FGNNATAN	149
DB	1106	FGNQAIN	1113

Search completed: August 2, 2004, 14:54:38
Job time : 29.7 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:35:42 ; Search time 44.9 Seconds
(without alignments)
950.215 Million cell updates/sec

Title: US-09-543-407-20
Perfect score: 774
Sequence: 1 MLLKVAFAAIVVSGSALA.....DSSVMVQVFGNNATNQY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04.*
1: Geneseq1980s.*
2: Geneseq1990s.*
3: Geneseq2000s.*
4: Geneseq2001s.*
5: Geneseq2002s.*
6: Geneseq2003as.*
7: Geneseq2003bs.*
8: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	774	100.0	151	3 AAB36350	Aab36350 Agfa::PT3
2	709	91.6	151	3 AAB36353	Aab36353 Agfa::PT3
3	689	89.0	151	2 AAR74625	Aar74625 Agfa sequ
4	689	89.0	151	3 AAB36341	Aab36341 Salmonell
5	684	88.4	151	2 AAW23570	Aaw23570 Salmonell
6	662	85.5	151	3 AAB36354	Aab36354 Agfa::PT3
7	622	80.4	151	3 AAB36349	Aab36349 Agfa::PT3
8	619	80.0	151	3 AAB36351	Aab36351 Agfa::PT3
9	611	78.9	151	3 AAB36346	Aab36346 Agfa::PT3
10	609	78.7	151	3 AAB36347	Aab36347 Agfa::PT3
11	606	78.3	151	3 AAB36352	Aab36352 Agfa::PT3
12	597	77.1	151	3 AAB36355	Aab36355 Agfa::PT3
13	574	74.2	151	3 AAB36348	Aab36348 Agfa::PT3
14	528	68.2	151	3 AAB36343	Aab36343 Escherich
15	523	67.6	151	7 ABR82651	Abr82651 E. coli C
16	504	65.1	120	2 AAR62761	Aar62761 Agfa sequ
17	504	65.1	120	2 AAW23569	Aaw23569 Salmonell
18	450	58.1	142	2 AAR52664	Aar52664 Fibronect
19	378	48.8	122	2 AAR52663	Aar52663 FNB curl1
20	153	19.8	45	3 AAB36316	Aab36316 Salmonell
21	132	17.1	22	3 AAB36318	Aab36318 Salmonell
22	123	15.9	23	3 AAB36321	Aab36321 Salmonell
23	123	15.9	23	3 AAB36326	Aab36326 Salmonell
24	123	15.9	23	3 AAB36338	Aab36338 Salmonell
25	113	14.6	24	7 ABR82644	Abr82644 E. coli C

26	111	14.3	22	3 AAB36322	Aab36322 Salmonell
27	111	14.3	22	3 AAB36327	Aab36327 Salmonell
28	111	14.3	22	3 AAB36337	Aab36337 Salmonell
29	109	14.1	23	3 AAB36340	Aab36340 Salmonell
30	109	14.1	23	3 AAB36324	Aab36324 Salmonell
31	109	14.1	23	3 AAB36319	Aab36319 Salmonell
32	102	13.2	26	7 ABR82649	Abr82649 E. coli V
33	96.5	12.5	151	3 AAB36344	Aab36344 Escherich
34	96	12.4	19	3 AAB36323	Aab36323 Salmonell
35	96	12.4	19	3 AAB36336	Aab36336 Salmonell
36	96	12.4	19	3 AAB36328	Aab36328 Salmonell
37	95	12.3	24	7 ABR82647	Abr82647 E. coli c
38	92	11.9	23	3 AAB36331	Aab36331 Escherich
39	91	11.8	186	6 ABU21488	Abu21488 Protein e
40	91	11.8	502	2 AAW32312	Aaw32312 Leishmani
41	90.5	11.7	677	4 AAG04318	Ag04318 Novel hum
42	90	11.6	24	7 ABR82642	Abr82642 E. coli N
43	89.5	11.6	423	4 AAG07164	Ag07164 Novel hum
44	89.5	11.6	447	3 AAG29728	Aag29728 Arabidops
45	89.5	11.6	468	3 AAG29727	Aag29727 Arabidops

ALIGNMENTS

RESULT 1

AAB36350
ID AAB36350 standard; protein; 151 AA.
XX
AC AAB36350;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa::PT3#5 amino acid sequence SEQ ID NO:20.
XX
KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127889P.
XX
(UYVI-) UNIV VICTORIA.
PI White AP, Doran JL, Collison SK, Kay WW;
XX
WPI: 2000-672631/65.
N-PSDB: AAC64826.

Recombinant agfa gene having a segment replaced by a foreign DNA sequence which encodes foreign epitope or antigen, expresses recombinant Agfa protein useful for eliciting immune response in animal.

Disclosure; Page 137; 139pp; English.

The present invention describes a recombinant agfa gene (I) where a segment of the gene has been replaced by a segment of a foreign DNA sequence which encodes a foreign epitope or antigen. Also described are: (1) use of thin aggregative fimbriae (SEF17/7AF) nucleation depended assembly system of strains of Salmonella, Escherichia coli and enterobacteriaceae for the production of fimbriae comprising recombinant Agfa, CSga and Agfa-homologue fimbriin subunits, respectively; (2) CC directing recombination of a recombinant gene into the chromosome of the homologous species; (3) directing recombination of a recombinant gene back into the chromosome of the homologous species, replacing the native

CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX Sequence 151 AA;
 SQ

Query Match 100.0%; Score 774; DB 3; Length 151;
 Best Local Similarity 100.0%; Pred. No. 8.8e-71;
 Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPWGGGNGHNGSGSPDSTLSIYQGSANAALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPWGGGNGHNGSGSPDSTLSIYQGSANAALQ 60

QY 61 SDARKYDQLVTRVVTHEMAHAGQGADNSTIETQNGFRNNATIDQWNAKNSDIIVGOYGG 120
 DB 61 SDARKYDQLVTRVVTHEMAHAGQGADNSTIETQNGFRNNATIDQWNAKNSDIIVGOYGG 120

QY 121 NNAALVNOTASDSSVWVRQVGFNNATANQY 151
 DB 121 NNAALVNOTASDSSVWVRQVGFNNATANQY 151

RESULT 2
 AAB36353
 ID AAB36353 standard; protein; 151 AA.
 AC AAB36353;
 XX
 XX 26-FEB-2001 (first entry)
 XX
 XX Agfa::PT3#8 amino acid sequence SEQ ID NO:26.
 DE Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 XX
 XX Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 XX WO20060102-A2.
 XX
 XX 12-OCT-2000.
 XX
 XX 05-APR-2000; 2000WO-CA000356.
 XX
 XX 05-APR-1999; 99US-0127888P.
 XX
 XX (UYVI-) UNIV VICTORIA.
 PA
 PI White AP, Doran JL, Collison SK, Kay WW;
 XX
 XX WPI; 2000-672631/65.
 DR N-PSDB; AAC64629.
 XX
 XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 XX which encodes foreign epitope or antigen, expresses recombinant Agfa
 XX protein useful for eliciting immune response in animal.
 XX
 XX Disclosure; Page 138; 139pp; English.

XX The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, Cga and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX Sequence 151 AA;
 SQ

Query Match 91.6%; Score 709; DB 3; Length 151;
 Best Local Similarity 89.9%; Pred. No. 3.7e-64;
 Matches 143; Conservative 0; Mismatches 0; Indels 16; Gaps 2;

QY 1 MKLLKVAFAAIVVSGSALAGVVPWGGGNGHNGSGSPDSTLSIYQGSANAALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPWGGGNGHNGSGSPDSTLSIYQGSANAALQ 57

QY 61 SDARKYDQLVTRVVTHEMAHA-----GQADNSTIETQNGFRNNATIDQWNAKNSD 112
 DB 58 -----YDQLVTRVVTHEMAHAGYNGADVGQADNSTIETQNGFRNNATIDQWNAKNSD 112

QY 113 ITVGQYGGNNAALVNOTASDSSVWVRQVGFNNATANQY 151
 DB 113 ITVGQYGGNNAALVNOTASDSSVWVRQVGFNNATANQY 151

RESULT 3
 AAR74625
 ID AAR74625 standard; protein; 151 AA.
 XX
 XX AAR74625;
 AC
 XX 25-MAR-2003 (revised)
 DT 26-JUN-1995 (first entry)
 XX
 XX Agfa sequence.
 XX
 XX Salmonella; Agfa; vaccine.
 KW
 XX Salmonella.
 OS
 XX WO9425598-A2.
 PN
 XX 10-NOV-1994.
 PD
 XX 26-APR-1994; 94WO-IB000207.
 PF
 XX 26-APR-1993; 93US-00054452.
 PR
 XX (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
 PA (KING/) KING J.
 XX

PI Kay WW, Collinson SK, Clouthier SC, Doran JL;
 XX WPI; 1994-358275/44.
 DR N-PSDB; AAQ87467.
 XX
 XX Eliciting an immune response to Salmonella - using attenuated Salmonella
 PT strains, vector constructs, or compans. contg. fimbrial type proteins.
 XX
 XX Disclosure; Fig 7B; 95pp; English.
 XX
 XX The Salmonella AgfA protein and DNA are used in vaccine and genetic
 CC immunization compositions, respectively, to elicit an immune response to
 CC Salmonella in animals (e.g. food producing animals) and humans. (Updated
 CC on 25-MAR-2003 to correct PN field.)
 XX
 XX Sequence 151 AA;
 SQ
 Query Match 89.0%; Score 689; DB 2; Length 151;
 Best Local Similarity 90.7%; Pred. No. 4.1e-62;
 Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
 QY 61 SDARKYDQLVTRVVTHEMAHAGQADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120
 DB 61 SDARKSETTITQSGYNGADVQGGADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120
 QY 121 NNAALVNTASDSSVMVRQVGFNNATANQY 151
 DB 121 NNAALVNTASDSSVMVRQVGFNNATANQY 151
 RESULT 4
 AAB36341 standard; protein; 151 AA.
 XX
 AC AAB36341;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Salmonella enteritidis AgfA amino acid sequence SEQ ID NO:5.
 XX
 KW Salmonella; agfA; chromosomal gene replacement; fimbrian; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 XX
 PN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 95US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collinson SK, Kay WW;
 DR WPI; 2000-672631/65.
 DR N-PSDB; AAC64617.
 XX
 XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant AgfA
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 135; 139pp; English.
 XX
 CC The present invention describes a recombinant agfA gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:

CC (1) use of thin aggregative fimbriae (SEF17/TAP) nucleation depended
 CC assembly system of strains of Salmonella. Escherichia coli and
 CC enterobacteriaceae for the production of fimbriae comprising recombinant
 CC AgfA, CsgA and AgfA-homologue fimbrian subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant AgfA
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrian protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 XX Sequence 151 AA;
 SQ
 Query Match 89.0%; Score 689; DB 3; Length 151;
 Best Local Similarity 90.7%; Pred. No. 4.1e-62;
 Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
 QY 61 SDARKYDQLVTRVVTHEMAHAGQADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120
 DB 61 SDARKSETTITQSGYNGADVQGGADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120
 QY 121 NNAALVNTASDSSVMVRQVGFNNATANQY 151
 DB 121 NNAALVNTASDSSVMVRQVGFNNATANQY 151
 RESULT 5
 AAW23570 standard; protein; 151 AA.
 ID AAW23570
 XX
 AC AAW23570;
 XX
 DT 25-MAR-2003 (revised)
 DT 29-SEP-1997 (first entry)
 XX
 DE Salmonella enteritidis 27655-3b agfA.
 XX
 KW Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody.
 XX
 OS Salmonella enteritidis.
 XX
 XX Key Location/Qualifiers
 FH Misc-difference 123
 FT /note= "Encoded by GCC"
 FT
 XX
 PN US5635617-A.
 XX
 PD 03-JUN-1997.
 XX
 XX 25-APR-1994; 94US-00233788.
 XX
 PR 26-APR-1993; 93US-00054452.
 XX
 XX (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
 XX Collinson SK, Kay WW, Doran JL;
 PI

XX WPI; 1997-309886/28.
 DR N-PSDB; AAT74142.
 XX
 XX Isolated Salmonella gene agfa - used for diagnosis of Salmonella or
 PT enteropathogenic bacteria of the Enterobacteriaceae family.
 XX
 XX Example 2; Fig 7; 85pp; English.
 XX
 CC The present sequence represents agfa encoded by the full agfa gene
 CC derived from Salmonella enteritidis 27655-3b. The nucleic acid can be
 CC used to provide diagnostic assays for Salmonella and/or enteropathogenic
 CC bacteria of the family Enterobacteriaceae. It can also be used to provide
 CC proteins and antibodies which can be used for assays. The nucleic acid
 CC sequence can be used to provide probes or primers which can specifically
 CC hybridise to nucleic acid molecules from greater than 99% of Salmonella
 CC strains that are pathogenic to warm-blooded animals relative to nucleic
 CC acid molecules from virtually all other microbial organisms. (Updated on
 CC 25-MAR-2003 to correct PF field.)
 XX
 XX Sequence 151 AA;
 SQ

Query Match 88.4%; Score 684; DB 2; Length 151;
 Best Local Similarity 90.1%; Pred. No. 1.3e-61;
 Matches 136; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIIWVSGSALAGVVPQWGGGNGHNGSGSDSTLSIYQGSANAALALQ 60
 DB 1 MKLLKVAFAAIIWVSGSALAGVVPQWGGGNGHNGSGSDSTLSIYQGSANAALALQ 60
 QY 61 SDARKYDQLVTRVVTHEMAHAGQGDNSTIETQNGFRNNATIDQWNAKNSDIITVGQYGG 120
 DB 61 SDARKSETTITQSGYNGADVQGGADNSTIETQNGFRNNATIDQWNAKNSDIITVGQYGG 120
 QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
 DB 121 NNPALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 6
 AAB36354
 ID AAB36354 standard; protein; 151 AA.
 AC AAB36354;
 XX
 XX 26-FEB-2001 (first entry)
 DT
 DE Agfa::PT3#9 amino acid sequence SEQ ID NO:28.
 XX
 KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 PN WO200060102-A2.
 XX
 XX 12-OCT-2000.
 XX
 XX 05-APR-2000; 2000WO-CA000356.
 XX
 XX 05-APR-1999; 99US-0127888P.
 XX
 XX (UYVI-) UNIV VICTORIA.
 XX
 XX White AP, Doran JL, Collison SK, Kay WW;
 XX
 DR WPI: 2000-672631/55.
 DR N-PSDB; AAC64630.
 XX
 XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa

protein useful for eliciting immune response in animal.
 Disclosure; Page 138; 139pp; English.
 The present invention describes a recombinant agfa gene (1) where a
 segment of the gene has been replaced by a segment of a foreign DNA
 sequence which encodes a foreign epitope or antigen. Also described are:
 (1) use of thin aggregative fimbriae (SEF17/7AF) nucleation depended
 assembly system of strains of Salmonella, Escherichia coli and
 Enterobacteriaceae for the production of fimbriae comprising recombinant
 Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 directing recombination of a recombinant gene into the chromosome of the
 homologous species; (3) directing recombination of a recombinant gene
 back into the chromosome of the homologous species, replacing the native
 copy of that gene; and (4) eliciting an immune response in an animal,
 comprising separating an amino acid polymer comprising a recombinant Agfa
 protein containing a replacement segment or segments of foreign amino
 acid sequence or sequences grown on a Salmonella, E. coli or
 Enterobacteriaceae host cell, from the host cell and introducing the
 polymer into the animal in conjunction with a carrier or diluent. (1) is
 useful for the expression of recombinant Agfa protein which is useful for
 eliciting an immune response in an animal. In a fimbrial presentation
 system the heterologous antigens are presented in high numbers (up to
 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 immunogenicity and adhesion properties relevant for an efficient live
 vaccine, the carrier fimbrial subunit proteins are usually strong
 immunogens, which may be important for directing an immune response
 against the inserted epitope, and hybrid fimbriae are easy and
 inexpensive to purify in large amount. The present sequence is given in
 the exemplification of the present invention

Sequence 151 AA;
 Query Match 85.5%; Score 662; DB 3; Length 151;
 Best Local Similarity 81.9%; Pred. No. 2.3e-59;
 Matches 136; Conservative 0; Mismatches 0; Indels 30; Gaps 2;

QY 1 MKLLKVAFAAIIWVSGSALAGVVPQWGGGNGHNGSGSDSTLSIYQGSANAALALQ 60
 DB 1 MKLLKVAFAAIIWVSGSALAGVVPQWGGGNGHNGSGSDSTLSIYQGSANAALALQ 60
 QY 61 SDARK-----YDQLVTRVVTHEMAHAGQGDNSTIETQNGFRNNATIDQ 105
 DB 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHA-----FRNNATIDQ 105
 QY 106 WNAKNSDIITVGQYGGNNAALVNQTASDSSVMVRQVGFNNATANQY 151
 DB 106 WNAKNSDIITVGQYGGNNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 7
 AAB36349
 ID AAB36349 standard; protein; 151 AA.
 AC AAB36349;
 XX
 XX 26-FEB-2001 (first entry)
 DT
 DE Agfa::PT3#4 amino acid sequence SEQ ID NO:18.
 XX
 KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 PN WO200060102-A2.
 XX
 XX 12-OCT-2000.
 XX
 XX 05-APR-2000; 2000WO-CA000356.
 XX
 XX


```

PR 05-APR-1999; 99US-0127888P.
PA (UYVI-) UNIV VICTORIA.
XX White AP, Doran JL, Collison SK, Kay WW;
XX WPI; 2000-672631/65.
DR N-PSDB; AAC64625.
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX Disclosure; Page 136; 139pp; English.
XX The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SFF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX Sequence 151 AA;
XX
XX Query Match 80.4%; Score 622; DB 3; Length 151;
XX Best Local Similarity 73.6%; Pred. No. 2.8e-55;
XX Matches 128; Conservative 0; Mismatches 0; Indels 46; Gaps 2;
XX
XX QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQYGSANAALALQ 60
XX Db 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQYGSANAALALQ 60
XX
XX QY 61 SDARKYDQLVTRVVTHEMAHA-----GQADNSTIELTQNGF 97
XX Db 43 -----YDQLVTRVVTHEMAHALQSDARKSETTITQSGYNGADVQGGADNSTIELTQNGF 97
XX
XX QY 98 RNNATIDOWNAKNSDITVGOYGGNNAALVNQTSADSSVMVQVGFGNNTANQY 151
XX Db 98 RNNATIDOWNAKNSDITVGOYGGNNAALVNQTSADSSVMVQVGFGNNTANQY 151
XX
XX RESULT 8
XX AAB36351
XX ID AAB36351 standard; protein; 151 AA.
XX AC AAB36351;
XX DT 26-FEB-2001 (first entry)
XX DE Agfa: PT3#6 amino acid sequence SEQ ID NO:22.
XX DX Salmonella; agfa; chromosomal gene replacement; fimbrial; epitope;
XX KW vaccine; immune response; immunogen.

```

```

XX Salmonella enteritidis.
XX Escherichia coli.
XX Synthetic.
XX WO200060102-A2.
XX 12-OCT-2000.
XX 05-APR-2000; 2000WO-CA000356.
XX 05-APR-1999; 99US-0127888P.
XX (UYVI-) UNIV VICTORIA.
XX White AP, Doran JL, Collison SK, Kay WW;
XX WPI; 2000-672631/65.
XX N-PSDB; AAC64627.
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX Disclosure; Page 137; 139pp; English.
XX The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SFF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX Sequence 151 AA;
XX
XX Query Match 80.0%; Score 619; DB 3; Length 151;
XX Best Local Similarity 74.6%; Pred. No. 5.6e-55;
XX Matches 129; Conservative 0; Mismatches 0; Indels 44; Gaps 2;
XX
XX QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQYGSANAALALQ 60
XX Db 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQYGSANAALALQ 60
XX
XX QY 61 SDARK-----YDQLVTRVVTHEMAHAGQADNSTIELTQNGF 98
XX Db 61 SDARKSETTITQSGYNGADVQGGADNDYDQLVTRVVTHEMAHA----- 103
XX
XX QY 99 RNNATIDOWNAKNSDITVGOYGGNNAALVNQTSADSSVMVQVGFGNNTANQY 151
XX Db 104 -----DOWNAKNSDITVGOYGGNNAALVNQTSADSSVMVQVGFGNNTANQY 151
XX
XX RESULT 9

```


CC the exemplification of the present invention
XX Sequence 151 AA;
SQ

Query Match 78.7%; Score 609; DB 3; Length 151;
Best Local Similarity 81.5%; Pred. No. 5.8e-54;
Matches 123; Conservative 6; Mismatches 22; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSSGPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSSGPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKYDQLVTRVTHEMAHAGQADNSTIELTQNGFRNNATIDOWNAKNSDITVGYGG 120
DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITVGYDQ 120
QY 121 NNAALVNOTASDVSVVRQVGFNNATANQY 151
DB 121 LVTRVTHEMAHAGVSVVRQVGFNNATANQY 151

RESULT 11
AAB36352
ID AAB36352 standard; protein; 151 AA.
XX
AC AAB36352;
DT 26-FEB-2001 (first entry)
DE AgfA::PT3#7 amino acid sequence SEQ ID NO:24.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
FN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UUVI-) UNIV VICTORIA.
PI White AP, Doran JL, Collison SK, Kay WW;
XX
XX WPI; 2000-672631/65.
XX
DR N-PSDB; AAC64628.
XX

Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 138; 139pp; English.

The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF1/7fAR) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or

CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 151 AA;
Query Match 78.3%; Score 606; DB 3; Length 151;
Best Local Similarity 82.1%; Pred. No. 1.2e-53;
Matches 124; Conservative 5; Mismatches 22; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSSGPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSSGPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKYDQLVTRVTHEMAHAGQADNSTIELTQNGFRNNATIDOWNAKNSDITVGYGG 120
DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITVTRVVT 120
QY 121 NNAALVNOTASDVSVVRQVGFNNATANQY 151
DB 121 HEMAHANQTASDVSVVRQVGFNNATANQY 151

RESULT 12
AAB36355
ID AAB36355 standard; protein; 151 AA.
XX
AC AAB36355;
DT 26-FEB-2001 (first entry)
DE AgfA::PT3#10 amino acid sequence SEQ ID NO:30.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
FN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UUVI-) UNIV VICTORIA.
PI White AP, Doran JL, Collison SK, Kay WW;
XX
XX WPI; 2000-672631/65.
XX
DR N-PSDB; AAC64631.
XX

Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 139; 139pp; English.

The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:

CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant AgfA
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX Sequence 151 AA;
 SQ
 Query Match 77.1%; Score 597; DB 3; Length 151;
 Best Local Similarity 81.5%; Pred. No. 9.8e-53;
 Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPWQGGGNGHNGSGDPSTLSIYQYGSANAALALQ 60
 Db 1 MKLLKVAFAAIVVSGSALAGVVPWQGGGNGHNGSGDPSTLSIYQYGSANAALALQ 60
 QY 61 SPARKYDQLVTRVVTHEMAHAGCGADNSTIETQNGFRNNATIDQWNAKNSDITVGQYGG 120
 Db 61 SPARKSETTITQSGYNGADVGQGDADNSTIETQNGFRNNATIDQWNAHAGG 120
 QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
 Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
 RESULT 13
 AAB36348
 ID AAB36348 standard; protein; 151 AA.
 XX
 AC AAB36348;
 XX
 DT 26-FEB-2001 (first entry)
 DE AgfA::PT3#3 amino acid sequence SEQ ID NO:16.
 XX Salmonella; agfA; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 PN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay WW;
 XX WPI; 2000-672631/65.
 DR

DR N-PSDB; AAC64624.
 XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant AgfA
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 136; 139pp; English.
 XX
 CC The present invention describes a recombinant agfA gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant AgfA
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX Sequence 151 AA;
 SQ
 Query Match 74.2%; Score 574; DB 3; Length 151;
 Best Local Similarity 80.8%; Pred. No. 2.2e-50;
 Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPWQGGGNGHNGSGDPSTLSIYQYGSANAALALQ 60
 Db 1 MKLLKVAFAAIVVSGSALAGVVPWQGGGNGHNGSGDPSTLSIYQYGSANAALALQ 60
 QY 61 SPARKYDQLVTRVVTHEMAHAGCGADNSTIETQNGFRNNATIDQWNAKNSDITVGQYGG 120
 Db 61 SPARKSETTITQSGYNGADVGQGDADNSTIETQNGFRNNATIDQWNAKNSDITVGQYGG 120
 QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
 Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
 RESULT 14
 AAB36343
 ID AAB36343 standard; protein; 151 AA.
 XX
 AC AAB36343;
 XX
 DT 26-FEB-2001 (first entry)
 DE Escherichia coli CsgA amino acid sequence SEQ ID NO:7.
 XX
 KW Salmonella; agfA; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Escherichia coli.
 OS Synthetic.
 XX
 PN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX

PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
XX WPI; 2000-672631/65.
DR N-PSDB; AAC64619.
XX
XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 135; 139pp; English.
XX
XX The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF1/7AF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
XX the exemplification of the present invention
SQ Sequence 151 AA;

Query Match 68.2%; Score 528; DB 3; Length 151;
Best Local Similarity 68.9%; Pred. No. 1.1e-45;
Matches 104; Conservative 20; Mismatches 27; Indels 0; Gaps 0;
QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHGGNSGSPDSTLSIYQYGSANAALAQ 60
Db 1 MKLLKVAATAIVFSGSALAGVVPQYGGGNGHGGNSGSPNSLNITVYGGNSALAQ 60
QY 61 SDARKYDQLVTRVVTHEMAHAGOGADNSTIELTQNGFRNNATIDOWNAKNSDITVGYGG 120
Db 61 TDARNSDLTITQHGCGNGADVGCGSDSSIDLITQRCFGNSATLDOWNKNSMTVKQFGG 120
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
Db 121 GNGAAVDQTASNSVNVTVQVGFNNATAHQY 151

RESULT 15
ABR82651
ID ABR82651 standard; protein; 151 AA.

XX ABR82651;

XX 04-DEC-2003 (first entry)

XX E. coli CsgA subunit 15 kDa protein.

KW Plasma protein; immune response; antibacterial; vaccine; gene therapy.
XX
OS Escherichia coli.
XX
PN WO2003064446-A2.
XX
PD 07-AUG-2003.
XX
XX 30-JAN-2003; 2003WO-EP000943.
PF
XX 31-JAN-2002; 2002GB-00002275.
PR
XX (HANS-) HANSA MEDICAL RES AB.
PA
XX Bjoerck L, Olsen A, Wikstroem M, Herwald H;
PI
XX WPI; 2003-846136/61.
DR N-PSDB; ACF36153.
XX
XX New isolated peptide capable of binding a mammalian plasma protein,
PT useful in the manufacture of a medicament for the prevention and/or
PT treatment of a bacterial infection, such as Escherichia coli, Salmonella
PT or Shigella infections.
XX
XX Disclosure; Page 41-42; 42pp; English.
XX
XX The invention relates to an isolated peptide capable of binding a
CC mammalian plasma protein or of generating an immune response in a mammal
CC selected from sequences shown in ABR82642, ABR82648-49. The peptide or
CC antibody is useful for treating a bacterial infection in a human or
CC animal or in the manufacture of a medicament for the prophylactic
CC treatment of a bacterial infection, such as Escherichia coli, Salmonella
CC or Shigella infection. The peptide that is immobilized on a solid support
CC is also useful as a reagent for determining the ability of a plasma
CC protein to bind to bacteria. The present sequence represents an E. coli
CC 15 kDa protein
XX
XX Sequence 151 AA;
SQ
Query Match 67.6%; Score 523; DB 7; Length 151;
Best Local Similarity 68.2%; Pred. No. 3.4e-45;
Matches 103; Conservative 20; Mismatches 28; Indels 0; Gaps 0;
QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHGGNSGSPDSTLSIYQYGSANAALAQ 60
Db 1 MKLLKVEATAIVFSGSALAGVVPQYGGGNGHGGNSGSPNSLNITVYGGNSALAQ 60
QY 61 SDARKYDQLVTRVVTHEMAHAGOGADNSTIELTQNGFRNNATIDOWNAKNSDITVGYGG 120
Db 61 TDARNSDLTITQHGCGNGADVGCGSDSSIDLITQRCFGNSATLDOWNKNSMTVKQFGG 120
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
Db 121 GNGAAVDQTASNSVNVTVQVGFNNATAHQY 151

Search completed: August 2, 2004, 14:48:26
Job time : 45.9 secs

This Page Blank (uspto)


```
; ORIGINAL SOURCE:
; ORGANISM: Pinctada fucata
; CELL TYPE: mantle epithelial cell
; FEATURE:
; NAME/KEY: peptide
; LOCATION: from 1 to 738
; IDENTIFICATION METHOD: E (by experiment)
US-08-864-038A-3

Query Match      10.8%; Score 83.5; DB 3; Length 738;
Best Local Similarity 27.7%; Pred. No. 2;
Matches 44; Conservative 12; Mismatches 56; Indels 47; Gaps 6;

QY 3 LKVAAPAAIIVSSALAGVVPQGGGHNHNGGNSGPDSTLSIYQGSNAALALQSD 62
DB 419 LKKSASASASASASAG-----CGGGGGNGGNGGGGG-----GGAGALAAALAA 465
QY 63 ARKYDQL-----VTRVVTHEMAHAGQADNSTIELTQNGFRNNATIDOWNAKNSDITVGOY 118
DB 466 AGAGGLGGGGGGGALAAALAAAGAGG-----GGFGGLGGL-----GGL 504
QY 119 GGNNAALVNQASDSS-----VMVRQVGFNNATA 148
DB 505 CGGSAALAAAAAASGGGGRALRRALRRMRGGGSA 543

RESULT 5
US-09-498-520A-18
; Sequence 18, Application US/09498520A
; Patent No. 6613553
; GENERAL INFORMATION:
; APPLICANT: Rock, Charles O
; APPLICANT: Heath, Richard J
; TITLE OF INVENTION: No. 6613553al Enoyl Reductases and Methods of Use Thereof
; FILE REFERENCE: SJ-0022
; CURRENT APPLICATION NUMBER: US/09/498,520A
; CURRENT FILING DATE: 2000-02-04
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 18
; LENGTH: 321
; TYPE: PRT
; ORGANISM: Caulobacter crescentus
US-09-498-520A-18

Query Match      10.3%; Score 79.5; DB 4; Length 321;
Best Local Similarity 24.5%; Pred. No. 1.7;
Matches 48; Conservative 15; Mismatches 64; Indels 69; Gaps 10;

QY 4 LKVAAPAAIIVVSGSA-----LAG---VVPQWGGGNGHNG----- 34
DB 109 LKAAGLKVMVYCGAVKHAQAEQAGCDAVICQGGEGGHTGLVGTPLVAQAEAVKIPV 168
QY 35 ---CGNSSGPDSTLSIYQGSANALALQSDARKYDQVTRVVTHEMAHAG-----Q 83
DB 169 VAAGGLHDG-----RGAA-AALALGAQG---VVMGTFRFIASHEAHAGDLYRQAVE 215
QY 84 GADNSTIEL-TQNG-----FRNNATIDOWNAKNSDITV-----GOYGGNNAALVN 127
DB 216 AADEDTVTRCYSGKPMKPVKKNPYVDWEARPGDIQPPQAMVSIIRGANGGIGGQIEG 275
QY 128 QTADSSVMVRQVGF 143
DB 276 LDKAKSCFANGQSAGG 291

RESULT 6
US-09-328-352-4764
; Sequence 4764, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Berton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
```

```
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4764
; LENGTH: 975
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-4764

Query Match      10.3%; Score 79.5; DB 4; Length 975;
Best Local Similarity 25.0%; Pred. No. 8;
Matches 34; Conservative 19; Mismatches 70; Indels 13; Gaps 4;

QY 15 SSGALAGVVPQGGGNGHNGGNSGPDSTLSIYQGSANALALQSDARKYDQVTRV 74
DB 279 NGTGDGSG-VSALGGSGSGDAGNGIASGNGEHNHYGNG---NGDDVDITAPITGVL 333
QY 75 THE-----MAHAGQADNSTIELTQNGFRNNATIDOWNAKNSDITVGOYGGNNAALVN 128
DB 334 NISGNSFTLIGNSSSSSVNTAPTTSTNTVNDNTID--NGNSGGTGGSGSGSGDGLNG 391
QY 129 TASDSSVMVRQVGF 144
DB 392 AASGNGEHNHYGNGN 407

RESULT 7
US-08-458-023B-6
; Sequence 6, Application US/08458023B
; Patent No. 5667990
; GENERAL INFORMATION:
; APPLICANT: Beika, Randy M.
; APPLICANT: Yoder, Wendy
; APPLICANT: Takagi, Shinobu
; APPLICANT: Boomathathan, Karuppan C.
; TITLE OF INVENTION: ASPERGILLUS EXPRESSION SYSTEM
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5667990o No. 5667990disk of No. 5667990th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10174-6201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/458,023B
; FILING DATE: 01-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lowney Dr., Karen A.
; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 4086.010-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 363 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-458-023B-6

Query Match      10.1%; Score 78; DB 1; Length 363;
Best Local Similarity 30.7%; Pred. No. 2.9;
Matches 35; Conservative 13; Mismatches 28; Indels 38; Gaps 8;
```



```
RESULT 11
US-09-823-494-26
; Sequence 26, Application US/09823494
; Patent No. 6355610
; GENERAL INFORMATION:
; APPLICANT: Chesebro, Bruce W
; APPLICANT: Caughey, Byron W
; APPLICANT: Chabry, Joelle
; APPLICANT: Priola, Susette
; TITLE OF INVENTION: Inhibitors of Formation of Protease Resistant Pion
; FILE REFERENCE: 50121
; CURRENT APPLICATION NUMBER: US/09/823,494
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/128,450
; PRIOR FILING DATE: 1998-08-03
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 26
; LENGTH: 254
; TYPE: PRT
; ORGANISM: Hamster sp.
US-09-823-494-26

Query Match          10.0%; Score 77.5; DB 4; Length 254;
Best Local Similarity 25.4%; Pred. No. 2;
Matches 32; Conservative 19; Mismatches 19; Indels 15; Gaps 6;

QY 26 WG-CGGNHGNGSGSDSTLSIYQGSANAALQSDARKYDOLVTRVTHEMAHAGQG 84
DB 89 WQGGGTHNQNKSKPTNNK--HMAGAAAAGAVVGGLOGY--MLGSAMSRPMHFGND 144
QY 85 ADNSTIELTQGFENNA---TIDQNAKNS-----DITVQYGGNNAAL-VNQTSADSS 134
DB 145 WEDRYRENWRYNQVYRVPDQYNNQNFVHDCVNITIKQHTVTTTGTGTFETDIK 204
QY 135 VMVQV 140
DB 205 IMERVV 210

RESULT 12
US-07-792-259-17
; Sequence 17, Application US/07792259
; Patent No. 5286638
; GENERAL INFORMATION:
; APPLICANT: TANAKA, YOSHIKAZU
; APPLICANT: ASHIKARI, TOSHIHIKO
; APPLICANT: HATANAKA, HARUYO
; APPLICANT: SHIBANO, YUJI
; APPLICANT: AMACHI, TERUO
; APPLICANT: NAKAYAMA, TORU
; APPLICANT: SUMIDA, MOTOO
; TITLE OF INVENTION: PEROXIDE GENE OF MICROBIAL ORIGIN
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN, DABBY & CUSHMAN
; STREET: 1615 L. STREET, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/792,259
; FILING DATE: 19911115
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SCOTT, WATSON T.

Query Match          9.9%; Score 77; DB 4; Length 943;
Best Local Similarity 26.2%; Pred. No. 14;
Matches 37; Conservative 14; Mismatches 66; Indels 24; Gaps 7;
```

```
; REGISTRATION NUMBER: 26581
; REFERENCE/DOCKET NUMBER: 9437/93433
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-861-3067
; TELEFAX: 202-822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 364 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-07-792-259-17

Query Match          10.0%; Score 77.5; DB 1; Length 364;
Best Local Similarity 30.4%; Pred. No. 3.3;
Matches 35; Conservative 13; Mismatches 28; Indels 39; Gaps 8;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGHNH---GGNSSG-----PDSTLS 46
DB 1 MKLSLFTFAAVIIGALAL---PQPGGGGSGVTCPGGQSTNSQCCVWFVDVLDLQTN 56
QY 47 IYQYGSANAALQSDARKYDOLVTRVTHE-----MAHAGQ-----GADNSTI 90
DB 57 FYQ-GS-----KCESPVRK---ILRIVFHAIGFSPALTAAGQFGGGGADGSI 101

RESULT 13
US-09-056-556-204
; Sequence 204, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 204:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 943 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-09-056-556-204

Query Match          9.9%; Score 77; DB 4; Length 943;
Best Local Similarity 26.2%; Pred. No. 14;
Matches 37; Conservative 14; Mismatches 66; Indels 24; Gaps 7;
```

TREAT

```
QY 16 GSALAGVVPQWGGG-GNHN-GGNSSGPDSITLSIYQGSANAA--LALQSDARKYDQLVT 71
|||
Db 464 GSGNIGVFNVGSGSLGNYNIGSGN-----LGIYNI GFNVGVDYVGFNAGDFNQGFA 516
|||
QY 72 RVVTHEMAHAGOGADNSTIELT---QNGFRNATIDQWNAKNSDITVQYGGNNAALVNQ 128
|||
Db 517 NTGNNNIGFANTGNNNIGIGLSDNQOGFN---IASGWSGTGNSGLFNSGTNNVGIFNA 573
|||
QY 129 TASDSSVMVRQVGFGNNTAN 149
|||
Db 574 GTGN-----VGIANSGTGN 587

RESULT 14
US-09-072-596-199
; Sequence 199, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neco, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Twardzik, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 199:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 943 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-09-072-596-199

Query Match 9.9%; Score 77; DB 4; Length 943;
Best Local Similarity 26.2%; Pred. No. 14;
Matches 37; Conservative 14; Mismatches 66; Indels 24; Gaps 7;

QY 16 GSALAGVVPQWGGG-GNHN-GGNSSGPDSITLSIYQGSANAA--LALQSDARKYDQLVT 71
|||
Db 464 GSGNIGVFNVGSGSLGNYNIGSGN-----LGIYNI GFNVGVDYVGFNAGDFNQGFA 516
|||
QY 72 RVVTHEMAHAGOGADNSTIELT---QNGFRNATIDQWNAKNSDITVQYGGNNAALVNQ 128
|||
Db 517 NTGNNNIGFANTGNNNIGIGLSDNQOGFN---IASGWSGTGNSGLFNSGTNNVGIFNA 573
|||
QY 129 TASDSSVMVRQVGFGNNTAN 149
|||
Db 574 GTGN-----VGIANSGTGN 587

US-09-072-596-199
; Sequence 199, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neco, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Twardzik, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 199:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 943 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-09-072-596-199

Query Match 9.9%; Score 77; DB 4; Length 943;
Best Local Similarity 26.2%; Pred. No. 14;
Matches 37; Conservative 14; Mismatches 66; Indels 24; Gaps 7;

QY 16 GSALAGVVPQWGGG-GNHN-GGNSSGPDSITLSIYQGSANAA--LALQSDARKYDQLVT 71
|||
Db 464 GSGNIGVFNVGSGSLGNYNIGSGN-----LGIYNI GFNVGVDYVGFNAGDFNQGFA 516
|||
QY 72 RVVTHEMAHAGOGADNSTIELT---QNGFRNATIDQWNAKNSDITVQYGGNNAALVNQ 128
|||
Db 517 NTGNNNIGFANTGNNNIGIGLSDNQOGFN---IASGWSGTGNSGLFNSGTNNVGIFNA 573
|||
QY 129 TASDSSVMVRQVGFGNNTAN 149
|||
Db 574 GTGN-----VGIANSGTGN 587
```

```
Db 574 GTGN-----VGIANSGTGN 587

RESULT 15
US-09-477-135A-131
; Sequence 131, Application US/09477135A
; Patent No. 6572865
; GENERAL INFORMATION:
; APPLICANT: Nano, Francis
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
; TITLE OF INVENTION: immunostimulatory Peptides
; FILE REFERENCE: 52888
; CURRENT APPLICATION NUMBER: US/09/477,135A
; CURRENT FILING DATE: 2000-01-03
; PRIOR APPLICATION NUMBER: 08990823
; PRIOR FILING DATE: 1997-12-15
; PRIOR APPLICATION NUMBER: US 96/10375
; PRIOR FILING DATE: 1996-06-14
; PRIOR APPLICATION NUMBER: 60/000,254
; PRIOR FILING DATE: 1995-06-15
; NUMBER OF SEQ ID NOS: 169
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 131
; TYPE: PRT
; LENGTH: 943
; ORGANISM: Mycobacterium tuberculosis
US-09-477-135A-131

Query Match 9.9%; Score 77; DB 4; Length 943;
Best Local Similarity 26.2%; Pred. No. 14;
Matches 37; Conservative 14; Mismatches 66; Indels 24; Gaps 7;

QY 16 GSALAGVVPQWGGG-GNHN-GGNSSGPDSITLSIYQGSANAA--LALQSDARKYDQLVT 71
|||
Db 571 GSGNIGVFNVGSGSLGNYNIGSGN-----LGIYNI GFNVGVDYVGFNAGDFNQGFA 623
|||
QY 72 RVVTHEMAHAGOGADNSTIELT---QNGFRNATIDQWNAKNSDITVQYGGNNAALVNQ 128
|||
Db 624 NTGNNNIGFANTGNNNIGIGLSDNQOGFN---IASGWSGTGNSGLFNSGTNNVGIFNA 680
|||
QY 129 TASDSSVMVRQVGFGNNTAN 149
|||
Db 681 GTGN-----VGIANSGTGN 694

Search completed: August 2, 2004, 14:58:34
Job time : 13 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:54:48 ; Search time 36.8 Seconds

(without alignments)
1287.123 Million cell updates/sec

Title: US-09-543-407-20

Perfect score: 774
Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVQVGFNNATANYQ 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	525	67.8	151	12	US-09-741-873B-4
2	525	67.8	151	12	Sequence 4, Appl1
3	447	57.8	131	12	Sequence 4, Appl1
4	447	57.8	131	12	Sequence 2, Appl1
5	101.5	13.1	445	15	Sequence 2, Appl1
6	94.5	12.2	438	14	Sequence 20638, A
7	91	11.8	186	12	Sequence 9343, A
8	85	11.0	3705	12	Sequence 49412, A
9	85	11.0	6310	12	Sequence 77944, A
10	84.5	10.9	1778	14	Sequence 67793, A
11	84	10.9	1129	12	Sequence 749, Appl
12	83.5	10.8	278	9	Sequence 48048, A
13	83	10.7	597	9	Sequence 28, Appl
14	83	10.7	678	12	Sequence 146, Appl
15	82.5	10.7	435	14	Sequence 64573, A
					Sequence 3213, Ap

Query Match 67.8%; Score 525; DB 12; Length 151;

Best Local Similarity 68.2%; Pred. No. 6e-46;

Matches 103; Conservative 21; Mismatches 27; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGSGSPDSTLSIVQYGSANAALAQ 60

ALIGNMENTS

RESULT 1

US-09-741-873B-4
; Sequence 4, Application US/09741873B
; Publication No. US20020081722A1
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; CURRENT FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-4

16 82.5 10.7 515 14 US-10-128-714-8213 Sequence 8213, Ap
17 82.5 10.7 2732 14 US-10-238-075-1119 Sequence 1119, Ap
18 82.5 10.7 2834 14 US-10-085-959-252 Sequence 252, App
19 82 10.6 65 9 US-09-996-194-16 Sequence 16, Appl
20 82 10.6 65 12 US-10-164-966-33 Sequence 33, Appl
21 82 10.6 354 10 US-09-820-843A-21 Sequence 21, Appl
22 82 10.6 440 12 US-10-424-599-229222 Sequence 22922, A
23 82 10.6 1862 12 US-10-282-122A-497577 Sequence 49757, A
24 81.5 10.5 2204 12 US-10-282-122A-64364 Sequence 64364, A
25 81 10.5 562 12 US-10-282-122A-64514 Sequence 64514, A
26 81 10.5 628 12 US-10-282-122A-53269 Sequence 53269, A
27 80.5 10.4 154 16 US-10-437-963-162284 Sequence 162284, A
28 80.5 10.4 486 15 US-10-369-493-20619 Sequence 20619, A
29 80 10.3 145 16 US-10-437-963-147748 Sequence 147748, A
30 80 10.3 477 12 US-10-425-114-70098 Sequence 70098, A
31 80 10.3 507 12 US-10-424-599-229226 Sequence 229226, A
32 80 10.3 538 12 US-10-425-114-68152 Sequence 68152, A
33 80 10.3 558 16 US-10-437-963-175203 Sequence 175203, A
34 79.5 10.3 562 14 US-10-156-761-13039 Sequence 13039, A
35 79.5 10.3 1721 12 US-10-282-122A-62548 Sequence 62548, A
36 79 10.2 209 12 US-10-424-599-921110 Sequence 221110, A
37 79 10.2 276 15 US-10-369-493-3641 Sequence 3641, Ap
38 79 10.2 688 14 US-10-032-585-7876 Sequence 7876, Ap
39 79 10.2 974 12 US-10-282-122A-44999 Sequence 44999, A
40 78.5 10.1 271 14 US-10-156-761-11721 Sequence 11721, A
41 78.5 10.1 292 16 US-10-437-963-195404 Sequence 195404, A
42 78.5 10.1 472 16 US-10-467-479-2 Sequence 18473, A
43 78 10.1 545 15 US-10-369-493-18473 Sequence 18473, A
44 78 10.1 594 14 US-10-156-761-13173 Sequence 13173, A
45 78 10.1 1846 12 US-10-282-122A-49773 Sequence 49773, A


```
Best Local Similarity 64.9%; Pred. No. 5.2e-38;
Matches 85; Conservative 20; Mismatches 26; Indels 0; Gaps 0;

Qy 21 GVVPQWGGGHNHGGSSGPDSTLSIYQGSANAALQSDARKYDQLVTRVVTHEMAH 80
Db 1 GVVPYQGGGHNHGGSSGNSLNTYQGGNSALALQTDARNSLTITQHGCGNGAD 60

Qy 81 AGQADNSTIELTQNGFRNNATIDOWNAKNSDITVQYGGNNAALVNOTASDSSVMYQV 140
Db 61 VQGSDDSSIDLTRQGFNSATLDQWNGKNSMTVKQFGGNGAAVQDTASNSSVNVTVQ 120

Qy 141 GFGNNATANQY 151
Db 121 GFGNNATAHQY 131

RESULT 5
US-10-369-493-20638
; Sequence 20638, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 20638
; LENGTH: 445
; TYPE: PRT
; ORGANISM: Rhodospseudomonas palustris
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(445)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-20638

Query Match 13.1%; Score 101.5; DB 15; Length 445;
Best Local Similarity 26.0%; Pred. No. 0.076;
Matches 45; Conservative 20; Mismatches 55; Indels 53; Gaps 7;

Qy 7 AAFAA-----IVVSGSALAGVVPQWGGG-----NHNG-----GNSSGPDSTLSIYQY 50
Db 19 AAFADGNTVYLNQTDNQANITQSGNGNSVGAFNGSGFLQENGTLGA-NLLTVKQS 77

Qy 51 GSANAALQSDARKYDQLVTRVVTHEMAHAGQA-----DNSTIELTQNGFRNNATID 104
Db 78 GNSNS-----VGRDIQKQSGAGNSAALFQGTGSDVLELQQTGTSNGAVPS 123

Qy 105 QWNAKN-----SDITVQYGGNNAALVNOTASDSSVMYQVQ 141
Db 124 GWNWTDGVPENKITQDSSSNGSKSVIQLQKKNVFIKQNGTNSVNVQIG 176

RESULT 6
US-10-156-761-9343
; Sequence 9343, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, NASHIRA
```

```
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 9343
; LENGTH: 438
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-9343

Query Match 12.2%; Score 94.5; DB 14; Length 438;
Best Local Similarity 25.9%; Pred. No. 0.39;
Matches 37; Conservative 27; Mismatches 46; Indels 33; Gaps 8;

Qy 1 MKLLKVAAPAAIVSGSALAGVVPQWGGGHNHGGSSGPDSTLSIYQGSANAALA-- 58
Db 1 MRSIRAAAVGAVTMSLALAAASAC---GGSGTGGGSDSP-KTLT-YWASNGGASIAVD 54

Qy 59 ---LQSDARKYDQ-----LVTRVVTHEMAHAGGADNSTIELTQN-GFRN 99
Db 55 KKVLPQLDKPBEQTGIRKVKLEWVPNSDLLNRLT--ATTSGGPDVNLNIGNTWSASLQA 112

Qy 100 NATIDOWNAKNSDITVQYGGNN 122
Db 113 TGALLPWDARNFD---KIGKXD 131

RESULT 7
US-10-282-122A-49412
; Sequence 49412, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carx, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
```



```
QY 10 AAIVVGSALAGVVPW-----GGGG-----NHGGGSSG---PDSTLSIYQGSAN 54
Db 111 AGIGTAGSAM---PTWAIAGLSILGVGAAADNNGGGSSGGSDPDTS-----156
QY 55 AALALQSDARKYDQVLTVRVTHEMAHAGCGADNSTIEL-----TONGFRNN 100
Db 157 -----APATPIDLVLS---PDGLRLTGRGEAGTTVAIRDAAGNLIGSGTGVGADGNFNV 207
QY 101 ATIDQWAKNSDITVGVGGNNAALVNQTSDDSSVMVRQVGFGNNTAN 149
Db 208 LNAQINSENLDVLTDAAGNVSPAGVATAPDATAPLAPTDLAINEQGN 256

RESULT 10
US-10-238-075-749
; Sequence 749, Application US/10238075
; Publication No. US20030148324A1
; GENERAL INFORMATION:
; APPLICANT: I.N.S.E.R.M.
; TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isolated
; FILE REFERENCE: BLANDINE
; CURRENT APPLICATION NUMBER: US/10/238,075
; PRIOR FILING DATE: 2002-09-10
; PRIOR APPLICATION NUMBER: 0003145
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 1576
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 749
; LENGTH: 1778
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-238-075-749

Query Match 10.9%; Score 84.5; DB 14; Length 1778;
Best Local Similarity 26.8%; Pred. No. 25;
Matches 40; Conservative 21; Mismatches 59; Indels 29; Gaps 7;

QY 5 KVAFAAIVVSGSLAGVVPW-----GGGNHNGGSSGPDSTLSIYQGSANALAL- 59
Db 906 KISSNSTDAINGSQLYGVADSFSTYLGCGADISDTGLSGPTTIGTGTDTNVGDALAI 965
QY 60 -----QSDARKYDQVLTVRVTHEMAHAGCGADNSTIELTQ--NGFRNNATIDQWAKN 110
Db 966 NTSFSTSLGDALLWD-----ATAGKFSAXHGINNAPSVITDVANGAVSSTSSDAINGSQ 1019

QY 111 ----SDITVGVGGNNAALVNQTSDDSSV 135
Db 1020 LYGVSDVIADALGN--AVVN---TDGSI 1043

RESULT 11
US-10-282-122A-48048
; Sequence 48048, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
```

```
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 48048
; LENGTH: 1129
; TYPE: PRT
; ORGANISM: Burkholderia cepacia
US-10-282-122A-48048

Query Match 10.9%; Score 84; DB 12; Length 1129;
Best Local Similarity 27.1%; Pred. No. 16;
Matches 46; Conservative 20; Mismatches 72; Indels 32; Gaps 7;

QY 8 AFAAIVVGSALAGVVP-----QWGGGNHN--GGNNSGPDSTLSIYQYG 51
Db 26 AQAGAAVRPGAHAGLSPRRIVMTLAAAYLGLFQAGAHQYAGGGSATGGASSISV--- 81
QY 52 SANAALQSDARKYDQVLTVRVTHEMAHAGCGADNSTIELTONGFRNNAT-IDQWAKN 110
Db 82 -GNGSVATOVNSTAFGNLSTAAGTSATA-LGPGAHAMGDGSTAVGINAQATGVDASLGV 139
QY 111 SDITVGVGGNNAALVNQTSDDSSVM-----VRQVGFGNNTAN-QY 151
Db 140 QAIGSGAYSVAIGNLSSATQSGAVMGSGAATGVAIGLGNNAFASQY 189

RESULT 12
US-09-810-264-28
; Sequence 28, Application US/09810264
; Patent No. US20020076775A1
; GENERAL INFORMATION:
; APPLICANT: Crane, Virginia C.
; APPLICANT: Famodu, Omolayo O.
; APPLICANT: Hu, Xu
; APPLICANT: Lu, Guihua
; APPLICANT: Zhang, Lingyu
; TITLE OF INVENTION: WRKY Transcription Factors and Methods
; FILE REFERENCE: Of Use
; CURRENT APPLICATION NUMBER: US/09/810,264
; CURRENT FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: US 60/190,467
; PRIOR FILING DATE: 2000-03-17
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Triticum aestivum
US-09-810-264-28

Query Match 10.8%; Score 83.5; DB 9; Length 278;
Best Local Similarity 20.8%; Pred. No. 3;
```



```
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3213
; LENGTH: 435
; TYPE: PRT
; ORGANISM: Aspergillus fumigatus
US-10-128-714-3213

Query Match      10.7%; Score 82.5; DB 14; Length 435;
Best Local Similarity 19.6%; Pred.No.6.6;
Matches 31; Conservative 30; Mismatches 56; Indels 31; Gaps 6;

Qy 13 VVSGSALAGVVPQGGGNNHNGGNSGPDSTLSIYQVGSANAALALQSDARKYDQ--- 69
Db 274 VLTGHKGSVTCVRNGGTGKI-----YTSSHDETIKIWNQAQSGSLLOTLSAHARVNHLS 329

Qy 70 ---VTRVVTHEMAHAGQ--GADNSTIELTQNGFR-----NNATIDQ-----WN-107
Db 330 TDFALRTAYHD--HTGKVPGSDTEKVAVKRFEQAAMVNNKIVEKLVSASDDFTMYLWD 387

Qy 108 AKNSDITVGOYGGNNAALVNOTASDSSVMVYQVGFNN 145
Db 388 PENSTKPIARLLGHQKEVNVHTFSPDMAYIASAGFDNH 425
```

Search completed: August 2, 2004, 15:36:11
Job time : 36.8 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:48:33 ; Search time 167.9 Seconds
(without alignments)
877,809 Million cell updates/sec

Title: US-09-543-407-20

Perfect score: 774

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVFGNNATANQY 151

Scoring table: BLOSUM62

Gapex 10.0 , Gapext 0.5

Searched: 6019581 seqs, 976053577 residues

Total number of hits satisfying chosen parameters: 6019581

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending_Patents_AA_Main.*

- 1: /cgn2_6/ptodata/2/paa/PCTUS_COMB.pcp.*
- 2: /cgn2_6/ptodata/2/paa/US06_COMB.pcp.*
- 3: /cgn2_6/ptodata/2/paa/US07_COMB.pcp.*
- 4: /cgn2_6/ptodata/2/paa/US080_COMB.pcp.*
- 5: /cgn2_6/ptodata/2/paa/US081_COMB.pcp.*
- 6: /cgn2_6/ptodata/2/paa/US082_COMB.pcp.*
- 7: /cgn2_6/ptodata/2/paa/US083_COMB.pcp.*
- 8: /cgn2_6/ptodata/2/paa/US084_COMB.pcp.*
- 9: /cgn2_6/ptodata/2/paa/US085_COMB.pcp.*
- 10: /cgn2_6/ptodata/2/paa/US086_COMB.pcp.*
- 11: /cgn2_6/ptodata/2/paa/US087_COMB.pcp.*
- 12: /cgn2_6/ptodata/2/paa/US088_COMB.pcp.*
- 13: /cgn2_6/ptodata/2/paa/US089_COMB.pcp.*
- 14: /cgn2_6/ptodata/2/paa/US090_COMB.pcp.*
- 15: /cgn2_6/ptodata/2/paa/US091_COMB.pcp.*
- 16: /cgn2_6/ptodata/2/paa/US092_COMB.pcp.*
- 17: /cgn2_6/ptodata/2/paa/US093_COMB.pcp.*
- 18: /cgn2_6/ptodata/2/paa/US094_COMB.pcp.*
- 19: /cgn2_6/ptodata/2/paa/US095_COMB.pcp.*
- 20: /cgn2_6/ptodata/2/paa/US096_COMB.pcp.*
- 21: /cgn2_6/ptodata/2/paa/US097A_COMB.pcp.*
- 22: /cgn2_6/ptodata/2/paa/US097B_COMB.pcp.*
- 23: /cgn2_6/ptodata/2/paa/US098_COMB.pcp.*
- 24: /cgn2_6/ptodata/2/paa/US099A_COMB.pcp.*
- 25: /cgn2_6/ptodata/2/paa/US099B_COMB.pcp.*
- 26: /cgn2_6/ptodata/2/paa/US100_COMB.pcp.*
- 27: /cgn2_6/ptodata/2/paa/US101_COMB.pcp.*
- 28: /cgn2_6/ptodata/2/paa/US102_COMB.pcp.*
- 29: /cgn2_6/ptodata/2/paa/US103_COMB.pcp.*
- 30: /cgn2_6/ptodata/2/paa/US104_COMB.pcp.*
- 31: /cgn2_6/ptodata/2/paa/US106_COMB.pcp.*
- 32: /cgn2_6/ptodata/2/paa/US107_COMB.pcp.*
- 33: /cgn2_6/ptodata/2/paa/US107_COMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
------------	-------	-------------	--------	----	-------------

1	774	100.0	151	19	US-09-543-407-20	Sequence 20, Appl
2	709	91.6	151	19	US-09-543-407-26	Sequence 26, Appl
3	689	89.0	151	19	US-09-543-407-5	Sequence 5, Appl
4	684	88.4	151	6	US-08-233-642A-57	Sequence 57, Appl
5	662	85.5	151	19	US-09-543-407-28	Sequence 28, Appl
6	622	80.4	151	19	US-09-543-407-18	Sequence 18, Appl
7	619	80.0	151	19	US-09-543-407-22	Sequence 22, Appl
8	611	78.9	151	19	US-09-543-407-12	Sequence 12, Appl
9	609	78.7	151	19	US-09-543-407-14	Sequence 14, Appl
10	606	78.3	151	19	US-09-543-407-24	Sequence 24, Appl
11	602	77.8	151	19	US-09-543-407-31	Sequence 31, Appl
12	597	77.1	151	19	US-09-543-407-30	Sequence 30, Appl
13	574	74.2	151	19	US-09-543-407-16	Sequence 16, Appl
14	528	68.2	151	19	US-09-543-407-7	Sequence 7, Appl
15	525	67.8	151	13	US-08-978-878-4	Sequence 4, Appl
16	525	67.8	151	21	US-09-741-873B-4	Sequence 4, Appl
17	523	67.6	151	33	US-60-352-946-2	Sequence 2, Appl
18	523	67.6	151	33	US-60-444-371-2	Sequence 2, Appl
19	504	65.1	120	6	US-08-233-642A-55	Sequence 55, Appl
20	470	60.7	109	19	US-09-543-407-34	Sequence 34, Appl
21	466	60.2	158	16	US-09-252-691C-5834	Sequence 5834, Ap
22	466	60.2	158	16	US-08-252-691C-5834	Sequence 5834, Ap
23	466	60.2	158	30	US-10-417-886-5834	Sequence 5834, Ap
24	447	57.8	131	13	US-08-978-878-2	Sequence 2, Appl
25	447	57.8	131	21	US-09-741-873B-2	Sequence 2, Appl
26	343	44.3	109	19	US-09-543-407-35	Sequence 35, Appl
27	263	34.0	68	19	US-09-543-407-37	Sequence 37, Appl
28	166.5	21.5	70	19	US-09-543-407-32	Sequence 32, Appl
29	153	19.8	48	19	US-09-543-407-39	Sequence 39, Appl
30	105.5	13.6	145	21	US-09-739-449-8854	Sequence 8854, Ap
31	105.5	13.6	145	23	US-09-803-110-8854	Sequence 8854, Ap
32	101.5	13.1	445	29	US-10-369-493-20638	Sequence 20638, A
33	101.5	13.1	445	33	US-60-360-039-20638	Sequence 20638, A
34	99	12.8	492	21	US-09-708-427-7657	Sequence 7657, Ap
35	99	12.8	544	21	US-09-708-427-7656	Sequence 7656, Ap
36	99	12.8	573	21	US-09-708-427-7655	Sequence 7655, Ap
37	99	12.8	955	24	US-09-935-625-7946	Sequence 7946, Ap
38	99	12.8	955	24	US-09-935-625-25158	Sequence 25158, A
39	99	12.8	1036	24	US-09-935-625-7945	Sequence 7945, Ap
40	99	12.8	1036	24	US-09-935-625-25157	Sequence 25157, A
41	99	12.8	1055	19	US-09-570-581A-1851	Sequence 1851, Ap
42	99	12.8	1055	19	US-09-573-655A-349	Sequence 349, App
43	99	12.8	1055	19	US-09-573-655A-667	Sequence 667, App
44	99	12.8	1055	19	US-09-573-655B-349	Sequence 349, App
45	99	12.8	1055	19	US-09-573-655B-667	Sequence 667, App

ALIGNMENTS

RESULT 1
US-09-543-407-20
; Sequence 20, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.

US-09-543-407-20

Query Match	100.0%	Score	774	DB	19	Length	151
Best Local Similarity	100.0%	Pred. No.	3.3e-75				
Matches	151	Conservative	0	Mismatches	0	Indels	0
Gaps	0						
Qy	1	MLKLKVAFAAALVSGSALAGVFWQGGGNGHNGSGSPDSTLSIYQYGSAALALQ	60				
Db	1	MLKLKVAFAAALVSGSALAGVFWQGGGNGHNGSGSPDSTLSIYQYGSAALALQ	60				
Qy	61	SDARKYDQLTRVVTHEVAHAGQAGADNSTIELTQNGFRNNAIDQWNKNSDITVGYQGG	120				
Db	61	SDARKYDQLTRVVTHEVAHAGQAGADNSTIELTQNGFRNNAIDQWNKNSDITVGYQGG	120				
Qy	121	NNAALVNQTSADSSVMVRFQVFGNNATANQY	151				
Db	121	NNAALVNQTSADSSVMVRFQVFGNNATANQY	151				

RESULT 2

```

US-09-543-407-26
; Sequence 26, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.486
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding P73 from GPe3 of Leishmania major.
US-09-543-407-26

```

Query Match	91.6%;	Score	709;	DB	19;	Length	151;
Best Local Similarity	89.9%;	Pred. No.	3.8e-68;				
Matches	143;	Conservative	0;	Mismatches	0;	Indels	16;
Gaps	2;						
Qy	1	MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGDPSTLSIYQGSANAALAQ	60				
Db	1	MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGDPSTLSIYQGSANAAL---	57				
Qy	61	SDARKYDQLVTRVVTHEMAHA-----CGQADNSTIELTQNGFNATTIDQNAKNSD	112				
Db	58	-----YDQLVTRVVTHEMAHGYCGADNVCGQADNSTIELTQNGFNATTIDQNAKNSD	112				
Qy	113	ITVGQYGGNNALVNOTASDSSVVMRVQVGFNNATANQY	151				
Db	113	ITVGQYGGNNALVNOTASDSSVVMRVQVGFNNATANQY	151				

RESULTS

```

/ SEQUENCE 5, APPLICATION US/09543407
/ GENERAL INFORMATION:
/ APPLICANT: White, Aaron P.
/ APPLICANT: Doran, James L.
/ APPLICANT: Collinson, S. Karen
/ APPLICANT: Kay, William W.
/ TITLE OF INVENTION: BACTERIAL FIBRILAR SYSTEM FOR
/ PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
/ FILE REFERENCE: 920043.406
/ CURRENT APPLICATION NUMBER: US/09/543,407

```

```

; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 151
; TYPE: prt
; ORGANISM: Salmonella enteritidis
US-09-543-407-5

```

Query Match	89.0%; Score 689; DB 19; Length 151;
Best Local Similarity	90.7%; Pred. No. 5.6e-66;
Matches 137; Conservative	3; Mismatches 11; Indels 0; Gaps 0;
Qy	1 MKLKVAAFAAIVVSGSALAGVVPQGGGNNHNGGNSSGPDSTLSIYQYSNAALALQ 60
Dd	1 MKLKVAAFAAIVVSGSALAGVVPQGGGNNHNGGNSSGPDSTLSIYQYSNAALALQ 60
Qy	61 SDARKVDQLVTRVVTHEMAHAGACDANSTIELTQNGFRNATIDWNAKSIDITVGQYGG 120 :
Dd	61 SDARKEITTQSGYNGADVGOGADNSTIELTQNGFRNATIDWNAKSIDITVGQYGG 120 :
Qy	121 NNAALVNQTASDSVMVRQVGFNNATANQY 151
Dd	121 NNAALVNCTASDSVMVRQVGFNNATANQY 151

RESULT 4

```

US-08-233-642A-57
; Sequence 57, Application US/08233642A
; GENERAL INFORMATION:
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Clouthier, Sharon C.
; APPLICANT: Doran, James L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
; TITLE OF INVENTION: BASED VACCINES
; NUMBER OF SEQUENCES: 58
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,642A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO. 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 151 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-233-642A-57

```

Query Match 88.4%; Score 684; DB 6; Length 151;
Best Local Similarity 90.1%; Pred. No. 1.9e-65;
Matches 136: Conservative 3; Mismatches 12; Indels 0; Gaps 0;

```

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKYDQLVTRVVTHEMAHAGAGDNSTIELTQNGFRNNATIDQWNAKNSDITVQYGG 120
Db 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGG 120
QY 121 NNAALVNQATSDSSVMVRQVGFNNATANQY 151
Db 121 NNPAALVNQATSDSSVMVRQVGFNNATANQY 151

RESULT 5
US-09-543-407-28
; Sequence 28, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-28

Query Match 85.5%; Score 662; DB 19; Length 151;
Best Local Similarity 81.9%; Pred. No. 4.7e-63; Indels 30; Gaps 2;
Matches 136; Conservative 0; Mismatches 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
QY 61 SDARK-----YDQLVTRVVTHEMAHAGQADNSTIELTQNGFRNNATIDQ 105
Db 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHA-----FRNNATIDQ 105
QY 106 WNAKNSDITVQYGGNNAALVNQATSDSSVMVRQVGFNNATANQY 151
Db 106 WNAKNSDITVQYGGNNAALVNQATSDSSVMVRQVGFNNATANQY 151

RESULT 6
US-09-543-407-18
; Sequence 18, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-18

Query Match 80.4%; Score 622; DB 19; Length 151;
Best Local Similarity 73.6%; Pred. No. 1e-58; Indels 46; Gaps 2;
Matches 128; Conservative 0; Mismatches 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKYDQLVTRVVTHEMAHA-----GQADNSTIELTQNGF 97
Db 43 -----YDQLVTRVVTHEMAHALQSDARKSETTITQSGYNGADYQGGADNSTIELTQNGF 97
QY 98 RNNATIDOWNAKNSDITVQYGGNNAALVNQATSDSSVMVRQVGFNNATANQY 151
Db 98 RNNATIDOWNAKNSDITVQYGGNNAALVNQATSDSSVMVRQVGFNNATANQY 151

RESULT 7
US-09-543-407-22
; Sequence 22, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 22
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-22

Query Match 80.0%; Score 619; DB 19; Length 151;
Best Local Similarity 74.6%; Pred. No. 2.2e-58; Indels 44; Gaps 2;
Matches 129; Conservative 0; Mismatches 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
QY 61 SDARK-----YDQLVTRVVTHEMAHAGQADNSTIELTQNGFR 98
Db 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHA----- 103
QY 99 RNNATIDOWNAKNSDITVQYGGNNAALVNQATSDSSVMVRQVGFNNATANQY 151
Db 104 -----DOWNAKNSDITVQYGGNNAALVNQATSDSSVMVRQVGFNNATANQY 151

RESULT 8
US-09-543-407-12
; Sequence 12, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.

```

```

; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-12

Query Match      78.9%; Score 611; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 1.6e-57;
Matches 122; Conservative 7; Mismatches 22; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQGSANAALALQ 60
QY 61 SDARKYDQLVTRVVTHEMAHAGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
Db 61 SDARKSETTITQSGYNGADVQGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
QY 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151
Db 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151

RESULT 9
US-09-543-407-14
; Sequence 14, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-14

Query Match      78.7%; Score 609; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 2.7e-57;
Matches 123; Conservative 6; Mismatches 22; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQGSANAALALQ 60
QY 61 SDARKYDQLVTRVVTHEMAHAGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
Db 61 SDARKSETTITQSGYNGADVQGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGYDD 120

```

```

QY 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151
Db 121 LVTRVVTHEMAHASVMVRQVGFNNATANQY 151

RESULT 10
US-09-543-407-24
; Sequence 24, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-24

Query Match      78.3%; Score 606; DB 19; Length 151;
Best Local Similarity 82.1%; Pred. No. 5.7e-57;
Matches 124; Conservative 5; Mismatches 22; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQGSANAALALQ 60
QY 61 SDARKYDQLVTRVVTHEMAHAGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
Db 61 SDARKSETTITQSGYNGADVQGGADNSTIELTQNGFRNNATIDQWNAKNDQLVTRVVT 120
QY 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151
Db 121 HEMAHANQTASDSSVMVRQVGFNNATANQY 151

RESULT 11
US-09-543-407-31
; Sequence 31, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Salmonella enteritidis
US-09-543-407-31

Query Match      77.8%; Score 602; DB 19; Length 131;
Best Local Similarity 89.3%; Pred. No. 1.3e-56;
Matches 117; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 21 GVVPQWGGGNGHNGSGSPDSTLSIYQGSANAALALQSDARKYDQLVTRVVTHEMAH 80

```


Db 1 GVFPWGGGNGHNGSGSPDSTLSIYQGSANAALALOSDARKSETTTTQSGYNGAD 60
QY 81 AGQADNSTIETQNGFRNNATIDQWNAKNSDITVQYGGNNAALVNQTASDSSVMVRQV 140
Db 61 VQCADNSTIETQNGFRNNATIDQWNAKNSDITVQYGGNNAALVNQTASDSSVMVRQV 120
QY 141 GFGNNATANQY 151
Db 121 GFGNNATANQY 131

RESULT 12

US-09-543-407-30
; Sequence 30, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543.407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-30

Query Match 77.1%; Score 597; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 5.4e-56;
Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQGSANAALALQ 60
QY 61 SDARKYDQLVTRVVTHEMAHAGADNSTIETQNGFRNNATIDQWNAKNSDITVQYGG 120
Db 61 SDARKSETTTTQSGYNGADVGQCADNSTIETQNGFRNNATYDQLVTRVVTHEMAHAG 120
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 13

US-09-543-407-16
; Sequence 16, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543.407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-16

Query Match 74.2%; Score 574; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 1.7e-53;
Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVYDQLVTRVVTHEMAHAGSPDSTLSIYQGSANAALALQ 60
QY 61 SDARKYDQLVTRVVTHEMAHAGADNSTIETQNGFRNNATIDQWNAKNSDITVQYGG 120
Db 61 SDARKSETTTTQSGYNGADVGQCADNSTIETQNGFRNNATIDQWNAKNSDITVQYGG 120
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 14

US-09-543-407-7
; Sequence 7, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543.407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-543-407-7

Query Match 68.2%; Score 528; DB 19; Length 151;
Best Local Similarity 68.9%; Pred. No. 1.6e-48;
Matches 104; Conservative 20; Mismatches 27; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQYGGGNGHNGSGSPDSTLSIYQGGNSALALQ 60
QY 61 SDARKYDQLVTRVVTHEMAHAGQCADNSTIETQNGFRNNATIDQWNAKNSDITVQYGG 120
Db 61 TDARNSDLTITQHGNGGADVGQSDSDSIDLTQRFGNSATLDQWNGKNSMTVKQFG 120
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
Db 121 GNGAAVDQTASNSVNVTVQVGFNNATANQY 151

RESULT 15

US-08-978-878-4
; Sequence 4, Application US/08978878
; GENERAL INFORMATION:
; APPLICANT: NORMARK, Staffan
; APPLICANT: OLSEN, Arne
; TITLE OF INVENTION: FIBRONECTIN BINDING PROTEIN AS WELL AS ITS PREPARATION
; FILE REFERENCE: 012889-081
; CURRENT APPLICATION NUMBER: US/08/978,878
; CURRENT FILING DATE: 1997-11-26
; EARLIER APPLICATION NUMBER: SE 8801723-1
; EARLIER FILING DATE: 1988-05-06

Tue Aug 3 10:54:37 2004

```

; EARLIER APPLICATION NUMBER: US 07/347,189
; EARLIER FILING DATE: 1989-05-04
; EARLIER APPLICATION NUMBER: US 07/789,437
; EARLIER FILING DATE: 1991-11-06
; EARLIER APPLICATION NUMBER: US 07/970,846
; EARLIER FILING DATE: 1992-11-03
; EARLIER APPLICATION NUMBER: US 08/187,865
; EARLIER FILING DATE: 1994-01-28
; EARLIER APPLICATION NUMBER: US 08/318,519
; EARLIER FILING DATE: 1994-10-05
; EARLIER APPLICATION NUMBER: US 08/495,959
; EARLIER FILING DATE: 1995-06-28
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-08-978-878-4

Query Match      67.8%; Score 525; DB 13; Length 151;
Best Local Similarity 68.2%; Pred. No. 3.5e-48;
Matches 103; Conservative 21; Mismatches 27; Indels 0; Gaps 0;

QY      1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQYGSANAALALQ 60
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      1 MKLLKVAAPAAIVFSSAVAGVVPQYGGGNGHGGGNSGPNSEINLYQYGGNSALALQ 60
      :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||

QY      61 SPARKYDQLVTRVTHEVAHAGQAGADNSTIBLTQNGFENNATIDOWNAKNSDITVQYGG 120
      :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
Db      61 TDARNSDLTITQHGNGGADVGQGGSDSSIDLTRQFGNSATLDQWNGKNSMTVTKQFGG 120
      :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||

QY      121 NNAALVNQTSADSSVMVRQVGFGNATANQY 151
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      121 GNGAAVDQTSANSSVNVTVQVGFGNATAHQY 151
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

Search completed: August 2, 2004, 15:26:44
Job time : 168.9 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:49:38 ; Search time 17.8 Seconds
(without alignments)
888.146 Million cell updates/sec

Title: US-09-543-407-20
Perfect score: 774
Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVGFNNATANY 151

Scoring table: ELOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 601315 seqs, 104695340 residues

Total number of hits satisfying chosen parameters: 601315

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents AA New:*

1: /cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pep:*

2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep:*

3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep:*

4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep:*

5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:*

6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:*

7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	525	67.8	151	5	US-09-741-873C-4
2	447	57.8	131	5	US-09-741-873C-2
3	87	11.2	573	7	US-60-565-632-7907
4	87	11.0	573	7	US-60-579-062-7907
5	85	11.0	147	6	US-10-425-115-193207
6	85	11.0	412	7	US-60-565-632-7905
7	85	11.0	412	7	US-60-579-062-7905
8	85	11.0	841	7	US-60-565-632-7906
9	85	11.0	841	7	US-60-579-062-7906
10	81.5	10.5	436	7	US-60-581-351-7636
11	80.5	10.4	511	6	US-10-425-115-320950
12	80	10.3	581	1	PCT-US04-07412-1695
13	80	10.3	581	6	US-10-393-559-1695
14	79.5	10.3	321	7	US-60-556-841-11319
15	79.5	10.3	321	7	US-60-566-841-11319
16	79	10.2	374	1	PCT-US04-11210-37
17	79	10.2	719	5	US-09-248-796A-17559
18	78.5	10.1	753	6	US-10-170-205E-35514
19	78.5	10.1	753	6	US-10-170-205E-35515
20	77	9.9	545	7	US-60-565-632-7973
21	77	9.9	545	7	US-60-579-062-7973
22	76.5	9.9	179	6	US-10-767-701-35342
23	76.5	9.9	254	1	PCT-US04-16242-10
24	76	9.8	556	6	US-10-425-115-337674
25	76	9.8	892	5	US-09-952-267B-5
26	76	9.8	892	6	US-10-872-768-5

27 76 9.8 892 6 US-10-872-768-5 Sequence 5, Appli

28 76 9.8 1095 1 PCT-US03-24982A-273 Sequence 273, App

29 76 9.8 1459 6 US-10-482-706-129 Sequence 129, App

30 75.5 9.8 167 6 US-10-425-115-242035 Sequence 242035,

31 75.5 9.8 276 6 US-10-425-115-339993 Sequence 339993,

32 75.5 9.8 415 5 US-09-490-324-280 Sequence 280, App

33 75.5 9.8 544 7 US-60-563-858-16 Sequence 16, Appl

34 75.5 9.8 619 1 PCT-US04-10229-38 Sequence 38, Appl

35 75 9.7 389 7 US-60-556-841-11867 Sequence 11867, A

36 75 9.7 552 5 US-09-248-796A-25337 Sequence 25337, A

37 74.5 9.6 179 6 US-10-425-115-346132 Sequence 346132,

38 74.5 9.6 298 7 US-60-565-632-8923 Sequence 8923, Ap

39 74.5 9.6 298 7 US-60-579-062-8923 Sequence 8923, Ap

40 74.5 9.6 634 7 US-60-581-351-9813 Sequence 9813, Ap

41 74.5 9.6 672 1 PCT-US04-07096-553 Sequence 553, App

42 74 9.6 279 6 US-10-425-115-33835 Sequence 33835,

43 74 9.6 358 5 US-09-248-796A-22578 Sequence 22578, A

44 74 9.6 389 7 US-60-556-841-9640 Sequence 9640, Ap

45 74 9.6 400 6 US-10-490-953-13 Sequence 13, Appl

ALIGNMENTS

RESULT 1

US-09-741-873C-4

; Sequence 4, Application US/09741873C

; GENERAL INFORMATION: Staffan

; APPLICANT: Normark, Staffan

; APPLICANT: Olsen, Arne

; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation

; FILE REFERENCE: 012889-084

; CURRENT APPLICATION NUMBER: US/09/741,873C

; CURRENT FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: SE 8801723-1

; PRIOR FILING DATE: 1998-05-06

; PRIOR APPLICATION NUMBER: US 08/978,878

; PRIOR FILING DATE: 1997-11-26

; PRIOR APPLICATION NUMBER: US 07/347,189

; PRIOR FILING DATE: 1989-05-04

; PRIOR APPLICATION NUMBER: US 07/789,437

; PRIOR FILING DATE: 1991-11-06

; PRIOR APPLICATION NUMBER: US 07/970,846

; PRIOR FILING DATE: 1992-11-03

; PRIOR APPLICATION NUMBER: US 08/187,865

; PRIOR FILING DATE: 1994-01-28

; PRIOR APPLICATION NUMBER: US 08/318,519

; PRIOR FILING DATE: 1994-10-05

; NUMBER OF SEQ ID NOS: 11

; SOFTWARE: Patent in version 3.0

; SEQ ID NO 4

; LENGTH: 151

; TYPE: PRT

; ORGANISM: Escherichia coli

US-09-741-873C-4

Query Match 67.8%; Score 525, DB 5; Length 151;

Best Local Similarity 68.2%; Pred. No. 8.3e-40;

Matches 103; Conservative 21; Mismatches 27; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGGNGGNSGSDPSTLSIYQVGSANAALQ 60

Db 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGGNGGNSGSDPSTLSIYQVGSANAALQ 60

QY 61 SDARKYDQLVTRVYVTHMAHAGQADNSTIELTQGNFRNATIDQWNAKNSDITVQCYGG 120

Db 61 TDARNSDLTITQHGCGGADVGQSDSDSLTQRCFGNSATLDQWNGKNSMTVQKFGG 120

QY 121 NNAALVQNTASDSSVMVRQVGFNNATANY 151

Db 121 GNGAAVDQTASNSSVNVTVQVGFNNATANY 151

RESULT 2
US-09-741-873C-2
; Sequence 2, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8601723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873C-2

Query Match 57.8%; Score 447; DB 5; Length 131;
Best Local Similarity 64.9%; Pred. No. 7,1e-33;
Matches 85; Conservative 20; Mismatches 26; Indels 0; Gaps 0;
QY 21 GVVPWGGGNGHNGSGPDSLSIYQGSANAALQSDARKYDQLVTRVVTHEMAH 80
Db 1 GVVPWGGGNGHNGSGPDSLSIYQGSANAALQSDARKYDQLVTRVVTHEMAH 80
QY 81 AGQADNSTIELTQNGFNATIDQNAKNSDITVQYGGNNAALVNOTASDSSVWVROV 140
Db 61 VQGGSDSSIDLTQRFNGSATLDQNGKNSMTVKQFGGNGAADVOTASNSVNTQV 120
QY 141 GFGNNATANOY 151
Db 121 GFGNNATANOY 131

RESULT 3
US-60-565-632-7907
; Sequence 7907, Application US/60565632
; GENERAL INFORMATION:
; APPLICANT: Monsanto Technology, LLC
; APPLICANT: Baum, James A
; APPLICANT: Kovalic, David K.
; APPLICANT: Larosa, Thomas J.
; APPLICANT: Lu, Maolong
; APPLICANT: Munyikwa, Tichifa R. I.
; APPLICANT: Roberts, James K.
; APPLICANT: Wu, Wei
; APPLICANT: Zhang, Bei
; TITLE OF INVENTION: Methods for Genetic Control of Insect Infestations in Plants and
; FILE REFERENCE: 38-21(53403)B
; CURRENT APPLICATION NUMBER: US/60/565,632
; CURRENT FILING DATE: 2004-04-27
; NUMBER OF SEQ ID NOS: 15449
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7907
; LENGTH: 573
; TYPE: PRT
; ORGANISM: Diabrotica virgifera

US-60-565-632-7907
Query Match 11.2%; Score 87; DB 7; Length 573;
Best Local Similarity 24.9%; Pred. No. 9.5;
Matches 42; Conservative 15; Mismatches 62; Indels 50; Gaps 7;
QY 30 GNHNGG--NSSGPDSTLSIYQGSANAALA-----LQSDARKYDQLVTRVVTHEMAH 80
Db 241 GNENGTGAENNAADAQTDVAQ--GSTNEANAENNAADVQNDAAQENGAAGAENSGNAD 299
QY 81 AGQADN-STIELTQ-----GFRNATIDOWN----- 107
Db 300 AAQGTDNGAAAEENTGNADPAQNDNGAAAEENGTAAENNAADVQNDAAQVNDNGA 359
QY 108 --AKNSDITVQYGGNNAALVNOTASD-----SSVMVROVFGNNATAN 149
Db 360 AAENNGNADAAQSDNDNGAAAEENNTNADQAQNGAAQSTANEANAENNAAD 408
RESULT 4
US-60-579-062-7907
; Sequence 7907, Application US/60579062
; GENERAL INFORMATION:
; APPLICANT: Baum, James A
; APPLICANT: Kovalic, David K
; APPLICANT: Larosa, Thomas J
; APPLICANT: Lu, Maolong
; APPLICANT: Munyikwa, Tichifa R. I.
; APPLICANT: Roberts, James K
; APPLICANT: Wu, Wei
; APPLICANT: Zhang, Bei
; TITLE OF INVENTION: Methods for Genetic Control of Insect Infestations in Plants and
; FILE REFERENCE: 38-21(53403) C
; CURRENT APPLICATION NUMBER: US/60/579,062
; CURRENT FILING DATE: 2004-06-11
; NUMBER OF SEQ ID NOS: 41445
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7907
; LENGTH: 573
; TYPE: PRT
; ORGANISM: Diabrotica virgifera
US-60-579-062-7907

Query Match 11.2%; Score 87; DB 7; Length 573;
Best Local Similarity 24.9%; Pred. No. 9.5;
Matches 42; Conservative 15; Mismatches 62; Indels 50; Gaps 7;
QY 30 GNHNGG--NSSGPDSTLSIYQGSANAALA-----LQSDARKYDQLVTRVVTHEMAH 80
Db 241 GNENGTGAENNAADAQTDVAQ--GSTNEANAENNAADVQNDAAQENGAAGAENSGNAD 299
QY 81 AGQADN-STIELTQ-----GFRNATIDOWN----- 107
Db 300 AAQGTDNGAAAEENTGNADPAQNDNGAAAEENGTAAENNAADVQNDAAQVNDNGA 359
QY 108 --AKNSDITVQYGGNNAALVNOTASD-----SSVMVROVFGNNATAN 149
Db 360 AAENNGNADAAQSDNDNGAAAEENNTNADQAQNGAAQSTANEANAENNAAD 408
RESULT 5
US-10-425-115-193207
; Sequence 193207, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115

```
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 193207
; LENGTH: 147
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MET4577_107790C.1.pep
US-10-425-115-193207

Query Match      11.0%; Score 85; DB 6; Length 147;
Best Local Similarity 24.6%; Pred. No. 2.8;
Matches 35; Conservative 18; Mismatches 63; Indels 26; Gaps 4;

QY 7 AAPAAIVVSGSALAGVVPWGCGGNHNGSGPDSSTLSIYQY-----GSANAALALQSD 62
Db 27 AQIPASATTGAATAATPQPGSGTGGVGVAGPDPAPLELYMEDIILGSSPRA----- 80
QY 63 ARKYDQLVTRVVTTHMAGAGQADNSTIELTQNGFRNNATIDOWNAKNSDITVGYQYGN 122
Db 81 -----RPIITGLGNIYNGQVPPFAPIGFSAPRNGVAIPNANGQ-----VPTYNGNT 126
QY 123 AALVNQRTASDSSVMVRQVGFN 144
Db 127 GIPLDTGLSRAGFL--QPTGQN 146

RESULT 6
US-60-565-632-7905
; Sequence 7905, Application US/60565632
; GENERAL INFORMATION:
; APPLICANT: Monsanto Technology, LLC
; APPLICANT: Baum, James A
; APPLICANT: Kovalic, David K.
; APPLICANT: Larosa, Thomas J.
; APPLICANT: Lu, Maolong
; APPLICANT: Munyikwa, Tichifa R. I.
; APPLICANT: Roberts, James K.
; APPLICANT: Wu, Wei
; APPLICANT: Zhang, Bei
; TITLE OF INVENTION: Methods for Genetic Control of Insect Infestations in Plants and
; FILE REFERENCE: 38-21(53403)B
; CURRENT FILING DATE: 2004-04-27
; NUMBER OF SEQ ID NOS: 15449
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7905
; LENGTH: 412
; TYPE: PRT
; ORGANISM: Diabrotica virgifera
; NAME/KEY: misc feature
; LOCATION: (369)..(369)
; OTHER INFORMATION: Xaa can be any naturally occurring amino acid
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (392)..(392)
; OTHER INFORMATION: Xaa can be any naturally occurring amino acid
; OTHER INFORMATION: Xaa can be any naturally occurring amino acid
US-60-579-062-7905

Query Match      11.0%; Score 85; DB 7; Length 412;
Best Local Similarity 26.4%; Pred. No. 9.6;
Matches 43; Conservative 12; Mismatches 62; Indels 46; Gaps 7;

QY 30 GNHNGGG--NSSGPDSTLSIYQYGSANAALA-----LQSDARKYDQLVTRVVTHEMAH 80
Db 85 GNENGTTGAENNAADAQTDAQ--GSTNEAENNAENNAADVQNDAAQANENGAAENGSGNAD 143
QY 81 AQCGADN--STIELTON-----GFRNATIDOWN----- 107
Db 144 AAQGTDNGAAAEENTGNADPAQGNNGGAAENGNGTAAENNAADVQNDAAQVNNNGA 203
QY 108 --AKNSDITVGOYGGNNAALVNQRTASDSSVMVRQVGFNATA 148
Db 204 AAENNGNADAQSNNDNGAAENNTNADAQNDAAQ--GTANEANA 245

RESULT 7
US-60-565-632-7906
; Sequence 7906, Application US/60565632
; GENERAL INFORMATION:
; APPLICANT: Monsanto Technology, LLC
; APPLICANT: Baum, James A
; APPLICANT: Kovalic, David K.
; APPLICANT: Larosa, Thomas J.
; APPLICANT: Lu, Maolong
; APPLICANT: Munyikwa, Tichifa R. I.
; APPLICANT: Roberts, James K.
; APPLICANT: Wu, Wei
; APPLICANT: Zhang, Bei
; TITLE OF INVENTION: Methods for Genetic Control of Insect Infestations in Plants and
; FILE REFERENCE: 38-21(53403)B
; CURRENT FILING DATE: 2004-04-27
; NUMBER OF SEQ ID NOS: 15449
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7905
; LENGTH: 412
; TYPE: PRT
; ORGANISM: Diabrotica virgifera
; NAME/KEY: misc feature
; LOCATION: (369)..(369)
; OTHER INFORMATION: Xaa can be any naturally occurring amino acid
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (392)..(392)
; OTHER INFORMATION: Xaa can be any naturally occurring amino acid
; OTHER INFORMATION: Xaa can be any naturally occurring amino acid
US-60-565-632-7905

Query Match      11.0%; Score 85; DB 7; Length 412;
Best Local Similarity 26.4%; Pred. No. 9.6;
Matches 43; Conservative 12; Mismatches 62; Indels 46; Gaps 7;

QY 30 GNHNGGG--NSSGPDSTLSIYQYGSANAALA-----LQSDARKYDQLVTRVVTHEMAH 80
Db 85 GNENGTTGAENNAADAQTDAQ--GSTNEAENNAENNAADVQNDAAQANENGAAENGSGNAD 143
QY 81 AQCGADN--STIELTON-----GFRNATIDOWN----- 107
Db 144 AAQGTDNGAAAEENTGNADPAQGNNGGAAENGNGTAAENNAADVQNDAAQVNNNGA 203
QY 108 --AKNSDITVGOYGGNNAALVNQRTASDSSVMVRQVGFNATA 148
Db 204 AAENNGNADAQSNNDNGAAENNTNADAQNDAAQ--GTANEANA 245
```

```
QY 108 --AKNSDITVGOYGGNNAALVNQRTASDSSVMVRQVGFNATA 148
Db 204 AAENNGNADAQSNNDNGAAENNTNADAQNDAAQ--GTANEANA 245

RESULT 7
US-60-579-062-7905
; Sequence 7905, Application US/60579062
; GENERAL INFORMATION:
; APPLICANT: Baum, James A
; APPLICANT: Kovalic, David K.
; APPLICANT: Larosa, Thomas J.
; APPLICANT: Lu, Maolong
; APPLICANT: Munyikwa, Tichifa R. I.
; APPLICANT: Roberts, James K.
; APPLICANT: Wu, Wei
; APPLICANT: Zhang, Bei
; TITLE OF INVENTION: Methods for Genetic Control of Insect Infestations in Plants and
; FILE REFERENCE: 38-21(53403)C
; CURRENT FILING DATE: 2004-06-11
; NUMBER OF SEQ ID NOS: 41445
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7905
; LENGTH: 412
; TYPE: PRT
; ORGANISM: Diabrotica virgifera
; NAME/KEY: misc feature
; LOCATION: (369)..(369)
; OTHER INFORMATION: Xaa can be any naturally occurring amino acid
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (392)..(392)
; OTHER INFORMATION: Xaa can be any naturally occurring amino acid
; OTHER INFORMATION: Xaa can be any naturally occurring amino acid
US-60-579-062-7905

Query Match      11.0%; Score 85; DB 7; Length 412;
Best Local Similarity 26.4%; Pred. No. 9.6;
Matches 43; Conservative 12; Mismatches 62; Indels 46; Gaps 7;

QY 30 GNHNGGG--NSSGPDSTLSIYQYGSANAALA-----LQSDARKYDQLVTRVVTHEMAH 80
Db 85 GNENGTTGAENNAADAQTDAQ--GSTNEAENNAENNAADVQNDAAQANENGAAENGSGNAD 143
QY 81 AQCGADN--STIELTON-----GFRNATIDOWN----- 107
Db 144 AAQGTDNGAAAEENTGNADPAQGNNGGAAENGNGTAAENNAADVQNDAAQVNNNGA 203
QY 108 --AKNSDITVGOYGGNNAALVNQRTASDSSVMVRQVGFNATA 148
Db 204 AAENNGNADAQSNNDNGAAENNTNADAQNDAAQ--GTANEANA 245

RESULT 8
US-60-565-632-7906
; Sequence 7906, Application US/60565632
; GENERAL INFORMATION:
; APPLICANT: Monsanto Technology, LLC
; APPLICANT: Baum, James A
; APPLICANT: Kovalic, David K.
; APPLICANT: Larosa, Thomas J.
; APPLICANT: Lu, Maolong
; APPLICANT: Munyikwa, Tichifa R. I.
; APPLICANT: Roberts, James K.
; APPLICANT: Wu, Wei
; APPLICANT: Zhang, Bei
; TITLE OF INVENTION: Methods for Genetic Control of Insect Infestations in Plants and
; FILE REFERENCE: 38-21(53403)B
; CURRENT FILING DATE: 2004-04-27
; NUMBER OF SEQ ID NOS: 15449
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7905
; LENGTH: 412
; TYPE: PRT
; ORGANISM: Diabrotica virgifera
; NAME/KEY: misc feature
; LOCATION: (369)..(369)
; OTHER INFORMATION: Xaa can be any naturally occurring amino acid
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (392)..(392)
; OTHER INFORMATION: Xaa can be any naturally occurring amino acid
; OTHER INFORMATION: Xaa can be any naturally occurring amino acid
US-60-565-632-7906
```

```
; NUMBER OF SEQ ID NOS: 15449
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 7906
; LENGTH: 841
; TYPE: PRT
; ORGANISM: Diabrotica virgifera
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (810)..(810)
; OTHER INFORMATION: Xaa can be any naturally occurring amino acid
US-60-565-632-7906

Query Match          11.0%; Score 85; DB 7; Length 841;
Best Local Similarity 30.7%; Pred. No. 23;
Matches 42; Conservative 13; Mismatches 56; Indels 26; Gaps 8;

QY 30 GNHNGGG--NSSGPPSTLSIYQGSANAALA-----LQSDARKYDQLVTRVVTHEMAH 80
Db 269 GNENGTCGAENNAADAQTDAQ--GSTNEANAENNANADVQNDAAQACANENGAAENSGNAD 327
QY 81 AQCGADN--STIELTON-----GFRNATIDQWNAKNSDITVQYGGNNAALVNQTSADSS 134
Db 328 AAQGTGNGAAAGTGNADPAQGNNGAA-----AENSGNENGTAAGNNA---NPDVQND 379
QY 135 VMVRQVGF--NNATAN 149
Db 380 AQVNDNGTAAENNGNAD 396

RESULT 9
US-60-579-062-7906
; Sequence 7906, Application US/60579062
; GENERAL INFORMATION:
; APPLICANT: Baum, James A
; APPLICANT: Kovalic, David K
; APPLICANT: Larosa, Thomas J
; APPLICANT: Lu, Maolong
; APPLICANT: Munyikwa, Tichifa R. I.
; APPLICANT: Roberts, James K
; APPLICANT: Wu, Wei
; APPLICANT: Zhang, Bei
; TITLE OF INVENTION: Methods for Genetic Control of Insect Infestations in Plants and
; FILE REFERENCE: Compositions thereof
; CURRENT APPLICATION NUMBER: US/60/579,062
; CURRENT FILING DATE: 2004-06-11
; NUMBER OF SEQ ID NOS: 41445
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 7906
; LENGTH: 841
; TYPE: PRT
; ORGANISM: Diabrotica virgifera
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (810)..(810)
; OTHER INFORMATION: Xaa can be any naturally occurring amino acid
US-60-579-062-7906

Query Match          11.0%; Score 85; DB 7; Length 841;
Best Local Similarity 30.7%; Pred. No. 23;
Matches 42; Conservative 13; Mismatches 56; Indels 26; Gaps 8;

QY 30 GNHNGGG--NSSGPPSTLSIYQGSANAALA-----LQSDARKYDQLVTRVVTHEMAH 80
Db 269 GNENGTCGAENNAADAQTDAQ--GSTNEANAENNANADVQNDAAQACANENGAAENSGNAD 327
QY 81 AQCGADN--STIELTON-----GFRNATIDQWNAKNSDITVQYGGNNAALVNQTSADSS 134
Db 328 AAQGTGNGAAAGTGNADPAQGNNGAA-----AENSGNENGTAAGNNA---NPDVQND 379
QY 135 VMVRQVGF--NNATAN 149
Db 380 AQVNDNGTAAENNGNAD 396

RESULT 10
US-60-579-902-7377
; Sequence 7377, Application US/60579902
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Krasomil-Osterfeld, Karina C.
; APPLICANT: Wu, Wei
; TITLE OF INVENTION: Nucleotide and Amino Acid Sequences from Xenorhabdus
; FILE REFERENCE: and Uses thereof
; CURRENT APPLICATION NUMBER: US/60/579,902
; CURRENT FILING DATE: 2004-06-15
; NUMBER OF SEQ ID NOS: 14985
; SEQ ID NO 7377
; LENGTH: 2663
; TYPE: PRT
; ORGANISM: Xenorhabdus bovienii
US-60-579-902-7377

Query Match          11.0%; Score 85; DB 7; Length 2663;
Best Local Similarity 24.7%; Pred. No. 91;
Matches 42; Conservative 23; Mismatches 59; Indels 46; Gaps 7;

QY 5 KVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGDPSTLSIYQGSANAALALQSDAR 64
Db 370 QLLASKSLVVSASAI-----DNROKITSSGGDITLTGQDLNLSGKIA---GQ 415
QY 65 KYDQLVTRVVTHEMAHAGQCADNSTIELTQNGFRNNA-----TIDQWNAKNS- 111
Db 416 HHQLQNTQSIINKEGH--MSADTVDTINTHQGLNNTAGLIVAERNMILFTGELLNRQGSV 473
QY 112 ----DITVQYQY-----GNNALVNQTSADSSVMVRQVFGNNATAN 149
Db 474 RSGQDLTNTGHNQDLNDRSLGQGLFSQGA-----MHLNTGYLNNQSGH 518

RESULT 11
US-60-581-351-7636
; Sequence 7636, Application US/60581351
; GENERAL INFORMATION:
; APPLICANT: Laurie, Cathy C
; TITLE OF INVENTION: Nucleic Acid Molecules Associated With Protein In Plants
; FILE REFERENCE: 38-21(53372)B
; CURRENT APPLICATION NUMBER: US/60/581,351
; CURRENT FILING DATE: 2004-06-17
; PRIOR APPLICATION NUMBER: US 60/479,962
; PRIOR FILING DATE: 2003-06-19
; NUMBER OF SEQ ID NOS: 13980
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 7636
; LENGTH: 436
; TYPE: PRT
; ORGANISM: Ceratopteris richardii
US-60-581-351-7636

Query Match          10.5%; Score 81.5; DB 7; Length 436;
Best Local Similarity 26.1%; Pred. No. 21;
Matches 29; Conservative 16; Mismatches 43; Indels 23; Gaps 3;

QY 29 GGNHNGGNS---SGPDSITLSIYQGSANAALALQSDARKYDQLVTRVVTHEMAHAGOGA 85
Db 77 GCGAGGSSVAGNAGTANINQOHSAEAGLLAIARS-----GGDLAQSGGG 127
QY 86 DNSTIELTQNGFRNATIDQWNAKNSDITVQYGGNNAALVNQTSADSSVM 136
Db 128 RGNLLDLHSD-----TANSSDLVEGHEGGHGGHNOQDSQVL 167

RESULT 12
US-10-425-115-320950
; Sequence 320950, Application US/10425115
```

```
/ GENERAL INFORMATION:
/ APPLICANT: La Rosa, Thomas J.
/ APPLICANT: Kovalic, David K.
/ APPLICANT: Zhou, Yihua
/ APPLICANT: Cao, Yongwei
/ TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
/ TITLE OF INVENTION: Plants
/ FILE REFERENCE: 38-2153222B
/ CURRENT APPLICATION NUMBER: US/10/425,115
/ CURRENT FILING DATE: 2003-04-28
/ NUMBER OF SEQ ID NOS: 369326
/ SEQ ID NO 320950
/ LENGTH: 511
/ TYPE: PRT
/ ORGANISM: Zea mays
/ FEATURE:
/ NAME/KEY: unsure
/ LOCATION: (1)..(511)
/ OTHER INFORMATION: unsure at all Xaa locations
/ FEATURE:
/ OTHER INFORMATION: Clone ID: MRT4577_55769C.1.pap
US-10-425-115-320950

Query Match          10.4%; Score 80.5; DB 6; Length 511;
Best Local Similarity 20.8%; Pred. No. 32;
Matches 38; Conservative 28; Mismatches 70; Indels 47; Gaps 6;

QY 1 MKLKVAFPAALVVS-----GSALGVVPQWGG-----GNHGGGNSGPD 42
DB 133 LRLLEERAQEAIALSCRDVLRASGRFRFEDAWAKVPGSDGVYAWVAANYALGRJGGDPN 192
QY 43 STLSTLYQVGSANAALALQSDA-----RKYDQLVTRVVTHEMAHAGQADNSTIE-- 91
DB 193 KTVGIIELGASAQLTFVSDEVLPKLSYNTFTGETTYLTNSFLNFGQNAQDSHEM 252
QY 92 -LTQNGFRNATID----QWNAKNSDITVGYGGNNAALVNQTASDSSVWVYQVGFNNA 146
DB 253 LRSRGSFXNGTILADPCAPRGYSRNEEMLRMSGASRSTLENQ-----YVANG 299
QY 147 TAN 149
DB 300 TGN 302

RESULT 13
PCT-US04-07412-1695
/ Sequence 1695, Application PC/TUS0407412
/ GENERAL INFORMATION:
/ APPLICANT: Tang, Y. Tom
/ APPLICANT: Wang, Jian-rui
/ APPLICANT: Zhang, Jie
/ APPLICANT: Ren, Feiyan
/ APPLICANT: Zhou, Ping
/ APPLICANT: Ma, Yunging
/ APPLICANT: Ghosh, Malabika
/ APPLICANT: Xue, Aidong J.
/ APPLICANT: Asundi, Vinod
/ APPLICANT: Zhao, Qing A.
/ APPLICANT: Wang, Dunrui
/ APPLICANT: Goodrich, Ryle W.
/ APPLICANT: Chen, Rui-hong
/ APPLICANT: Wehrman, Tom
/ APPLICANT: Wang, Gezhi
/ APPLICANT: Wang, Zhiwei
/ APPLICANT: Boyle, Bryan J.
/ APPLICANT: Drmanac, Radoje T.
/ TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
/ FILE REFERENCE: 822CIP/PCT
/ CURRENT APPLICATION NUMBER: PCT/US04/07412
/ CURRENT FILING DATE: 2004-03-19
/ PRIOR APPLICATION NUMBER: US 10/389,559
/ PRIOR FILING DATE: 2003-03-14
/ PRIOR APPLICATION NUMBER: US 60/365,264

Query Match          10.3%; Score 80; DB 1; Length 581;
Best Local Similarity 22.3%; Pred. No. 41;
Matches 27; Conservative 25; Mismatches 47; Indels 22; Gaps 5;

QY 45 LSTLYQV-GSANAALALQSDARKYDQL--VTRVVTHEMAHAGQADNSTIELTQNGFRNNA 101
DB 242 VSLYPLAGGATQAFAXENCKAYKETYGVSHIFRHDMLQIPKQQQNEKYQVPQ---FDQS 298
QY 102 TI-----DOWNAKNSDITVGYGGNNAALV-----NOTASDSSVWVYQVGFNNA 145
DB 299 TIKNIESAKGLDVMDSWPLQADGTVAEYNGYHVFPALAGSPKADDTSTYMFYQKVGDN 358
QY 146 A 146
DB 359 S 359

RESULT 14
US-10-389-559-1695
/ Sequence 1695, Application US/10389559
/ GENERAL INFORMATION:
/ APPLICANT: Tang, Y. Tom
/ APPLICANT: Wang, Jian-rui
/ APPLICANT: Zhang, Jie
/ APPLICANT: Ren, Feiyan
/ APPLICANT: Zhou, Ping
/ APPLICANT: Ma, Yunging
/ APPLICANT: Ghosh, Malabika
/ APPLICANT: Xue, Aidong J.
/ APPLICANT: Asundi, Vinod
/ APPLICANT: Zhao, Qing A.
/ APPLICANT: Wang, Dunrui
/ APPLICANT: Goodrich, Ryle W.
/ APPLICANT: Chen, Rui-hong
/ APPLICANT: Wehrman, Tom
/ APPLICANT: Wang, Gezhi
/ APPLICANT: Wang, Zhiwei
/ APPLICANT: Boyle, Bryan J.
/ APPLICANT: Drmanac, Radoje T.
/ TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
/ FILE REFERENCE: 822
/ CURRENT APPLICATION NUMBER: US/10/389,559
/ CURRENT FILING DATE: 2003-03-14
/ PRIOR APPLICATION NUMBER: US 60/365,264
/ PRIOR FILING DATE: 2002-03-14
/ PRIOR APPLICATION NUMBER: US 60/340,187
/ PRIOR FILING DATE: 2001-12-12
```

```
; PRIOR APPLICATION NUMBER: PCT/US00/35017
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: US 09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: PCT/US01/02623
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: US 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: PCT/US01/03800
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 09/496,914
; PRIOR FILING DATE: 2000-02-03
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1920
; SOFTWARE: pt.fl_genes Version 6.0
; SEQ ID NO 1695
; LENGTH: 581
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-389-559-1695

Query Match      10.3%; Score 80; DB 6; Length 581;
Best Local Similarity 22.3%; Pred. No. 41;
Matches 27; Conservative 25; Mismatches 47; Indels 22; Gaps 5;

QY 45 LSIYQY-GSANAALQSDARKYDQL--VTRVVTHEMAHAGQADNSTIELTONGFRNNA 101
Db 242 VSLYPLAGGATCAFKENNQKAYKETGVSHTRHDMQLQIPKQQQNEKYQVPQ---FDOS 298

QY 102 TI-----DOWNAKNSDITVQYGGNNAALV-----NOTASDSSVMVRQVGFNN 145
Db 299 TIKNIESAKGLDWSNPLQADGTVAEYNGYHVVFALAGSPKADDTISYMFYKVGDN 358

QY 146 A 146
Db 359 S 359

RESULT 15
US-60-556-841-11319
; Sequence 11319, Application US/60556841
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S.
; TITLE OF INVENTION: Genes and Uses for Plant Improvement
; FILE REFERENCE: 38-21(53450)
; CURRENT APPLICATION NUMBER: US/60/556,841
; CURRENT FILING DATE: 2004-03-25
; NUMBER OF SEQ ID NOS: 12463
; SEQ ID NO 11319
; LENGTH: 321
; TYPE: PRT
; ORGANISM: Caulobacter crescentus CB15
; US-60-556-841-11319

Query Match      10.3%; Score 79.5; DB 7; Length 321;
Best Local Similarity 24.5%; Pred. No. 22;
Matches 48; Conservative 15; Mismatches 64; Indels 69; Gaps 10;

QY 4 LKVAAPAAATVWSGSA-----LAG--VVPQWGGGNGHNG----- 34
Db 109 LKAAGLKNVAVGAVKHAVKACQAGCDAVICQGGGGGHTGLVTLPLVAQAVEAVKIPV 168

QY 35 ---GNNSSGPDSTLSYQYGSANAALQSDARKYDQLVTRVVTHEMAHAG-----Q 83
Db 169 VAAGGLHDG-----RGLA-AALALGAQG---VVMGTREFIASHEAHAGDLVRQAVVE 215

QY 84 GADNSTIEL-TONG---FRNNATIDOWNAKNSDITV-----QYGGNNAALVN 127
Db 216 AADEDTVTRCYSGKPMRVKKNPYVDDWEARPDIOFPFOQAMVSIIRNGAMGGIGGQIEG 275
```

```
Qy 128 QTASDSSVMVROYGFG 143
Db 276 LDKAKSCFAMQSQAGG 291
```

Search completed: August 2, 2004, 15:29:53
Job time : 18.8 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:39:53 ; Search time 9.4 Seconds

(without alignments)
1545.204 Million cell updates/sec

Title: US-09-543-407-20

Perfect score: 774

Sequence: 1 MKLLKVAAPAAIVVSGSALA.....DSSVMVRQVGFNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78.*

1: PIR1.*

2: PIR2.*

3: PIR3.*

4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	689	89.0	151	2 JC6039	fimbrin protein ag
2	689	89.0	151	2 A10635	major curlin chain
3	528	68.2	151	2 S70788	curlin protein csg
4	506.5	65.4	152	2 D90806	curlin major subun
5	506.5	65.4	152	2 H85665	hypothetical prote
6	105.5	13.6	145	2 A31143	conserved hypothet
7	105.5	13.6	145	2 H98144	hypothetical prote
8	99	12.8	573	2 C86266	F319.21 protein -
9	96.5	12.5	151	2 S70787	curlin nucleator p
10	96.5	12.5	151	2 C90806	minor curlin subun
11	96.5	12.5	151	2 G85665	curlin minor chain
12	96	12.4	2174	2 E95965	hypothetical glyci
13	92.5	12.0	1567	2 S11672	ice nucleation pro
14	91.5	11.8	1322	2 S07053	ice nucleation pro
15	91	11.8	602	1 P10221	leishmanolysin (EC
16	91	11.8	645	2 F70825	probable PPE prote
17	90	11.6	639	2 C42049	leishmanolysin (EC
18	89.5	11.6	374	2 T03875	probable homeobox
19	89.5	11.6	375	2 T03874	probable homeobox
20	89.5	11.6	447	2 G84687	probable disease r
21	89	11.5	652	2 E97857	cell surface antig
22	88.5	11.4	151	2 JC6040	fimbrin protein ag
23	88.5	11.4	151	2 AH0635	nucleation compone
24	88.5	11.4	1588	2 A96036	probable adhesin Z
25	88.5	11.4	1588	2 H91188	probable adhesin E
26	88	11.4	590	1 A45621	leishmanolysin (EC
27	88	11.4	599	2 B42049	leishmanolysin (EC
28	88	11.4	599	2 A44951	leishmanolysin (EC
29	88	11.4	646	1 S19916	leishmanolysin (EC

RESULT 1

JC6039
fimbrin protein agfa precursor - Salmonella enteritidis
C;Species: Salmonella enteritidis
C;Date: 31-Dec-1996 #sequence revision 31-Dec-1996 #text_change 08-Oct-1999
C;Accession: JC6039; PC6015; A44898
R;Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Baner, P.A.; Kay, W.W.
J. Bacteriol. 178, 662-667, 1996
A;Title: Salmonella enteritidis agfBAC operon encoding thin, aggregative fimbriae.
A;Reference number: JC6039; MUID:96146512; PMID:8550497
A;Accession: JC6039
A;Molecule type: DNA
A;Residues: 1-151 <COL>
A;Cross-references: GB:U43280; NID:gl184712; PIDN:AC43599.1; PID:gl184714
A;Accession: PC6015
A;Molecule type: protein
A;Residues: 21-52 <CO2>
A;Experimental source: strain 27855-3b
A;Note: the authors translated the codon ACG for residue 44 as Ile
R;Collinson, S.K.; Emody, L.; Muller, K.H.; Trust, T.J.; Kay, W.W.
J. Bacteriol. 173, 4773-4781, 1991
A;Title: Purification and characterization of thin, aggregative fimbriae from Salmonella
A;Reference number: A4898; MUID:91310586; PMID:1677357
A;Contents: 27655
A;Accession: A4898
A;Status: preliminary
A;Molecule type: protein
A;Residues: 21-33 <CO3>
A;Note: sequence extracted from NCBI backbone (NCBIP:45936)
C;Genetics:
A;Gene: agfa
C;Function:
A;Description: major component of thin aggregative fimbriae
A;Note: fimbriae bind to fibronectin, plasminogen, tissue plasminogen activator
C;Keywords: fimbria
F;1-20/Domain: signal sequence #status predicted <SIG>
F;21-151/Product: fimbrin protein agfa #status experimental <MAT>

Query Match 89.0%; Score 689; DB 2; Length 151;
Best Local Similarity 90.7%; Pred. No. 3.1e-51;
Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

Qy 1 MKLLKVAAPAAIVVSGSALAGVVPQGGGNNHNGGNSGPDSTLSIYQVGSANAALQ 60
Db 1 MKLLKVAAPAAIVVSGSALAGVVPQGGGNNHNGGNSGPDSTLSIYQVGSANAALQ 60
Qy 61 SDARKYDQLVTRVVTTHMAHAGQADNSTIETQNGFRNNATIDQWNAKNSDITVQYGG 120
Db 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATIDQWNAKNSDITVQYGG 120
Qy 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

protein kinase sgg
cnjB protein - Tet
hemolysin (importe
probable secreted
probable PPE prote
hypothetical prote
probable autotrans
F7-2 fimbrial prot
ice nucleation act
ice nucleation pro
hypothetical prote
hypothetical prote
probable PPE prote
probable PPE prote
CREB-binding prote

Db 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151

RESULT 2

AI0635

major curlin chain precursor [imported] - Salmonella enterica subsp. enterica serovar Typhimurium

C:Species: Salmonella enterica subsp. enterica serovar Typhi

A:Note: This species has also been called Salmonella typhi

C>Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002

C:Accession: AI0635

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.

A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhimurium

A:Reference number: AB0502; PMID:11677608

A:Accession: AI0635

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-151 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD08268.1; PID:gl6502315; GSPDB:GN00176

C:Genetics:

A:Gene: STY1181

Query Match 89.0%; Score 689; DB 2; Length 151;

Best Local Similarity 90.7%; Pred. No. 3.1e-51;

Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIVSGSALAGVVPQWGGGNGGNSGPDSTLSIYQYGSANAALALQ 60

DB 1 MKLLKVAAPAAIVSGSALAGVVPQWGGGNGGNSGPDSTLSIYQYGSANAALALQ 60

QY 61 SDARKYDQVLRVVTTHMAHAGQADNSTIELTQNGFRNATIDOWNAKNSDITVGOYGG 120

DB 61 SPARKSETITQSGYNGAGDVQGGADNSTIELTQNGFRNATIDOWNAKNSDITVGOYGG 120

QY 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151

DB 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151

RESULT 3

S70788

curlin protein csgA precursor - Escherichia coli (strain K-12)

N:Alternate names: csgA protein; major curlin protein

C:Species: Escherichia coli

C>Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 01-Mar-2002

C:Accession: S70788; G64846; S31202; S34560; S34559

R:Hammar, M.; Arngvist, A.; Bian, Z.; Olsen, A.; Normark, S.

Mol. Microbiol. 18, 661-670, 1995

A:Title: Expression of two csg operons is required for production of fibronectin- and curli

A:Reference number: S70783; PMID:96414468; PMID:8817489

A:Accession: S70788

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-151 <HAM>

A:Cross-references: EMBL:X90754; NID:gl147558; PIDN:CAA62282.1; PID:gl147564

A:Experimental source: strain K12, substrain W3110

A:Note: The nucleotide sequence was submitted to the EMBL Data Library, August 1995

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; PMID:97426617; PMID:9278503

A:Accession: G64846

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-151 <BLAT>

A:Cross-references: GB:AB000205; GB:U00096; NID:gl1787265; PIDN:AAC74126.1; PID:gl1787279;

A:Experimental source: strain K-12, substrain M6165

R:Olsen, A.; Arngvist, A.; Hammar, M.; Sukupolvi, S.; Normark, S.

Mol. Microbiol. 7, 523-536, 1993

A:Title: The Ppos sigma factor relieves H-NS-mediated transcriptional repression of csgA

A:Reference number: S31202; PMID:93211294; PMID:8459772

A:Accession: S31202

A:Molecule type: DNA

A:Residues: 1-6,'V', 8-151 <OLS1>

A:Cross-references: EMBL:L04979

A:Accession: S34560

A:Molecule type: protein

A:Residues: 21-42;44-50 <OLS2>

R:Olsen, A.N.; Arngvist, A.M.

submitted to the EMBL Data Library, October 1992

A:Reference number: S34559

A:Accession: S34559

A:Molecule type: DNA

A:Residues: 1-133,'RQRDSGMWU' <OLS3>

A:Cross-references: EMBL:L04979; NID:G290424; PIDN:AAA23616.1; PID:G290425

A:Experimental source: strain K-12, substrain W3110

C:Genetics:

A:Gene: csgA

A:Map position: 23.15

C:Function:

A:Description: major component of wild-type curli; interaction between CsgA and CsgB tri

A:Note: curli are thin, coiled fibers expressed on the surface of Escherichia coli that

and H-kininogen; in the absence of CsgA, CsgB can self-assemble into polymers

F:1-20/Domain: signal sequence #status predicted <SIG>

F:21-151/Product: curlin #status experimental <MAT>

Query Match 68.2%; Score 528; DB 2; Length 151;

Best Local Similarity 68.9%; Pred. No. 1.2e-37;

Matches 104; Conservative 20; Mismatches 27; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIVSGSALAGVVPQWGGGNGGNSGPDSTLSIYQYGSANAALALQ 60

DB 1 MKLLKVAAPAAIVSGSALAGVVPQWGGGNGGNSGPDSTLSIYQYGSANAALALQ 60

QY 61 SDARKYDQVLRVVTTHMAHAGQADNSTIELTQNGFRNATIDOWNAKNSDITVGOYGG 120

DB 61 TDARNSDLTITQHGNGGADVGQGGSDSDITLQTFGNSATLDQWNGKNSMTVQFGG 120

QY 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151

DB 121 NGGAADVQTASNSVNVTVQVGFNNATAHQY 151

RESULT 4

D90806

curlin major subunit CsgA [imported] - Escherichia coli (strain O157:H7, substrain R1MD

C:Species: Escherichia coli

C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001

C:Accession: D90806

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.

gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen

A:Reference number: A99629; PMID:21156231; PMID:11298796

A:Accession: D90806

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-152 <HAY>

A:Cross-references: GB:BA000007; PIDN:BA34843.1; PID:gl3360880; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain R1MD 0509952

C:Genetics:

A:Gene: ECS1420

Query Match 65.4%; Score 506.5; DB 2; Length 152;

Best Local Similarity 67.1%; Pred. No. 7.7e-36;

Matches 102; Conservative 21; Mismatches 26; Indels 1; Gaps 1;

QY 1 MKLLKVAAPAAIVSGSALAGVVPQWGGGNGGNSGPDSTLSIYQYGSANAALALQ 59

DB 1 MKLLKVAAPAAIVSGSALAGVVPQWGGGNGGNSGPDSTLSIYQYGSANAALALQ 60

QY 60 QSDARKYDQVLRVVTTHMAHAGQADNSTIELTQNGFRNATIDOWNAKNSDITVGOYGG 119

```

Db      61 QADARNSLDITIQHGGNGAGDVGGGSDSSIDLQRFNGSATLDQWNGKDSHTVTKQG 120
QY      120 GNNAAALVNQTASDSSVMVVRQVGFNNATANQY 151
Db      121 GNGAAVDQTASNTVNTVQVGFNNATAHQY 152

RESULT 5
H85665
hypothetical protein csgA [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
C/Species: Escherichia coli
C/Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C/Accession: H85665
R/Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoukis, K.; Apodaca,
Nature 409, 529-533, 2001
A/Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A/Reference number: A85480; MUID:21074935; PMID:11206551
A/Accession: H85665
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-152 <STO>
A/Cross-references: GB:AB005174; NID:g12514574; PIDN:AA655788.1; GSPDB:GN00145; UWGP:Z16
A/Experimental source: strain O157:H7, substrain EDL933
C/Genetics:
A/Map position: linear chromosome

Query Match      65.4%; Score 506.5; DB 2; Length 152;
Best Local Similarity 67.1%; Pred. No. 7.7e-36;
Matches 102; Conservative 21; Mismatches 28; Indels 1; Gaps 1;

QY      1 MLLKVAFAAIVVSGSALAGVVPQW-GGGNGHGGNSGPDSTLSIYQYGSANAAL 59
Db      1 MLLKVAFAAIVVSGSALAGVVPQYGGGNGHGGNSGPNSELNIYQYGGNSALAL 60
QY      60 QSDARKYDQLVTRVVTHEMAHAGAGADNSTIELTQNGFRNNATIDOWNAKNSDITVQY 119
Db      61 QADARNSLDITIQHGGNGAGDVGGGSDSSIDLQRFNGSATLDQWNGKDSHTVTKQG 120
QY      120 GNNAAALVNQTASDSSVMVVRQVGFNNATANQY 151
Db      121 GNGAAVDQTASNTVNTVQVGFNNATAHQY 152

RESULT 6
AD3143
conserved hypothetical protein Atu4768 [imported] - Agrobacterium tumefaciens (strain C58)
C/Species: Agrobacterium tumefaciens
C/Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C/Accession: AD3143
R/Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.;
Karp, P.; Romero, P.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClellan
Science 294, 2317-2323, 2001
A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A/Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A/Reference number: AB2577; MUID:21608550; PMID:11743193
A/Accession: AD3143
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-145 <KUR>
A/Cross-references: GB:AB008689; PIDN:AAL45562.1; PID:g17743277; GSPDB:GN00187
A/Experimental source: strain C58 (Dupont)
C/Genetics:
A/Gene: Atu4768
A/Map position: linear chromosome

Query Match      13.6%; Score 105.5; DB 2; Length 145;
Best Local Similarity 23.2%; Pred. No. 0.049;
Matches 36; Conservative 27; Mismatches 61; Indels 31; Gaps 4;

Db      61 QADARNSLDITIQHGGNGAGDVGGGSDSSIDLQRFNGSATLDQWNGKDSHTVTKQG 120
QY      120 GNNAAALVNQTASDSSVMVVRQVGFNNATANQY 151
Db      121 GNGAAVDQTASNTVNTVQVGFNNATAHQY 152

RESULT 7
H98144
hypothetical protein AGR_L_228 [imported] - Agrobacterium tumefaciens (strain C58, Cere
C/Species: Agrobacterium tumefaciens
C/Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
C/Accession: H98144
R/Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.
Science 294, 2323-2328, 2001
A/Title: Genome sequence of the plant pathogen and Biotechnology Agent Agrobacterium tu
A/Reference number: A97359; MUID:21608551; PMID:11743194
A/Accession: H98144
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-145 <KUR>
A/Cross-references: GB:AE007870; PIDN:AAK8682.1; PID:g15158413; GSPDB:GN00170
C/Genetics:
A/Gene: AGR_L_228
A/Map position: linear chromosome

Query Match      13.6%; Score 105.5; DB 2; Length 145;
Best Local Similarity 23.2%; Pred. No. 0.049;
Matches 36; Conservative 27; Mismatches 61; Indels 31; Gaps 4;

QY      3 LKVAFAAIVVSGSALAGVVPQWGG-----GNHNGGNSGPDSTLSIYQYGSANA 55
Db      1 MIRKSFIALVALVGLSAAAFAMANDVRIEQYGSNSAGAGQGYGNRIYQNGYNR 60
QY      56 ALALQSDARKYDQLVTRVVTHEMAHAGAGADNSTIELTQNGFRNNATIDOWNAKNSDITV 115
Db      61 IVGHQ-----YGR-----HNL SAVQEGHDNYGSTTQNGNRN VAGI----- 96

QY      116 GOYGGNNAALVNQTASDSSVMVVRQVGFNNATANQ 150
Db      97 GQFGSNHTILTQDGNNGNIAAGVQVGRGCSANVSQ 131

RESULT 8
C86266
F319.21 protein - Arabidopsis thaliana
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Nov-2001
C/Accession: C86266
R/Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso
Chain, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Hawley, L.; Huizar, L.
Nature 408, 816-820, 2000
A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lures, J.S.; Maiti, R.; Marziali
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A/Reference number: A86141; MUID:21016719; PMID:11130712
A/Accession: C86266
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-573 <STO>
A/Cross-references: GB:AE005172; NID:g4850402; PIDN:AAD31072.1; GSPDB:GN00141
C/Genetics:

```

```

QY      3 LKVAFAAIVVSGSALAGVVPQWGG-----GNHNGGNSGPDSTLSIYQYGSANA 55
Db      1 MIRKSFIALVALVGLSAAAFAMANDVRIEQYGSNSAGAGQGYGNRIYQNGYNR 60
QY      56 ALALQSDARKYDQLVTRVVTHEMAHAGAGADNSTIELTQNGFRNNATIDOWNAKNSDITV 115
Db      61 IVGHQ-----YGR-----HNL SAVQEGHDNYGSTTQNGNRN VAGI----- 96

QY      116 GOYGGNNAALVNQTASDSSVMVVRQVGFNNATANQ 150
Db      97 GQFGSNHTILTQDGNNGNIAAGVQVGRGCSANVSQ 131

RESULT 7
H98144
hypothetical protein AGR_L_228 [imported] - Agrobacterium tumefaciens (strain C58, Cere
C/Species: Agrobacterium tumefaciens
C/Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
C/Accession: H98144
R/Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.
Science 294, 2323-2328, 2001
A/Title: Genome sequence of the plant pathogen and Biotechnology Agent Agrobacterium tu
A/Reference number: A97359; MUID:21608551; PMID:11743194
A/Accession: H98144
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-145 <KUR>
A/Cross-references: GB:AE007870; PIDN:AAK8682.1; PID:g15158413; GSPDB:GN00170
C/Genetics:
A/Gene: AGR_L_228
A/Map position: linear chromosome

Query Match      13.6%; Score 105.5; DB 2; Length 145;
Best Local Similarity 23.2%; Pred. No. 0.049;
Matches 36; Conservative 27; Mismatches 61; Indels 31; Gaps 4;

QY      3 LKVAFAAIVVSGSALAGVVPQWGG-----GNHNGGNSGPDSTLSIYQYGSANA 55
Db      1 MIRKSFIALVALVGLSAAAFAMANDVRIEQYGSNSAGAGQGYGNRIYQNGYNR 60
QY      56 ALALQSDARKYDQLVTRVVTHEMAHAGAGADNSTIELTQNGFRNNATIDOWNAKNSDITV 115
Db      61 IVGHQ-----YGR-----HNL SAVQEGHDNYGSTTQNGNRN VAGI----- 96

QY      116 GOYGGNNAALVNQTASDSSVMVVRQVGFNNATANQ 150
Db      97 GQFGSNHTILTQDGNNGNIAAGVQVGRGCSANVSQ 131

RESULT 8
C86266
F319.21 protein - Arabidopsis thaliana
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Nov-2001
C/Accession: C86266
R/Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso
Chain, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Hawley, L.; Huizar, L.
Nature 408, 816-820, 2000
A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lures, J.S.; Maiti, R.; Marziali
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A/Reference number: A86141; MUID:21016719; PMID:11130712
A/Accession: C86266
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-573 <STO>
A/Cross-references: GB:AE005172; NID:g4850402; PIDN:AAD31072.1; GSPDB:GN00141
C/Genetics:

```

A;Map position: 1

Query Match 12.8%; Score 99; DB 2; Length 573;
 Best Local Similarity 25.4%; Pred. No. 0.81; Mismatches 24; Conservative 34; Gaps 6;
 Matches 34; Conservative 24; Mismatches 24; Indels 30; Gaps 6;
 QY 13 VVSGALAGVWPQWGGGNGHNGSGPDTLSIYQGSANAALALQSDARKYDQVLVTR 72
 DB 84 IVSGGTVEG---KYRNDGGHNG---ISGPDTRSDVYPOASSGFGAKGLNID-----127
 QY 73 VVTHEMAHAGQADNSTIELTQNGRNNATIDOWNAKNS-----DITVGOYGGNN 122
 DB 128 IQSNKIAQGG---STTVVNNHGFSGNAVNPEVPHNSGAPPGQAQIIPVSQMSVNP 183
 QY 123 AALVNOTASDSSVM 136
 DB 184 NVMMKNSPTQSPVV 197

RESULT 9
 S70787
 curlin nucleator protein csbB precursor - Escherichia coli (strain K-12)
 N;Alternate names: csbB protein; curlin nucleation component; minor curlin protein
 C;Species: Escherichia coli
 C;Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 01-Mar-2002
 C;Accession: S70787; F64846
 R;Hammar, M.; Arngqvist, A.; Bian, Z.; Olsen, A.; Normark, S.
 Mol. Microbiol. 18, 661-670, 1995
 A;Title: Expression of two csb operons is required for production of fibronectin- and Cc
 A;Reference number: S70783; MUID:96414468; PMID:8817489
 A;Accession: S70787
 A;Status: nucleic acid sequence not shown; translation not shown
 A;Molecule type: DNA
 A;Residues: 1-151 <HAM>
 A;Cross-references: EMBL:X90754; NID:G1147558; PIDN:CAA62281.1; PID:G1147563
 A;Experimental source: strain K12, substrain W3110
 A;Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995
 R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
 A.; Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A;Title: The complete genome sequence of Escherichia coli K-12.
 A;Reference number: A64720; MUID:97426617; PMID:9278503
 A;Accession: F64846
 A;Status: nucleic acid sequence not shown; translation not shown
 A;Molecule type: DNA
 A;Residues: 1-151 <BLAT>
 A;Cross-references: GB:AE000205; GB:U00096; NID:G1787265; PIDN:AAAC74125.1; PID:G1787278;
 A;Experimental source: strain K-12, substrain MG1655
 C;Genetics:
 A;Gene: csbB

A;Map position: 23.15
 C;Function:
 A;Description: minor component of wild-type curlin; interaction between CsgA and CsgB tri
 A;Note: curlin are thin, coiled fibers expressed on the surface of Escherichia coli that
 and H-kinogen; in the absence of CsgA, CsgB can self-assemble into polymers
 F;1-21/Domain: signal sequence #status predicted <SIG>
 F;22-151/Product: minor curlin chain #status predicted <MAT>

Query Match 12.5%; Score 96.5; DB 2; Length 151;
 Best Local Similarity 28.1%; Pred. No. 0.29; Mismatches 15; Conservative 32; Indels 5; Gaps 2;
 Matches 32; Conservative 15; Mismatches 62; Indels 5; Gaps 2;

QY 38 SSGPDTLSIYQGSANAALALQSDARKYDQVLVTRVTHEMAHAGQADNSTIELTQNGF 97
 DB 21 AAGYDLANSEYNF----AVNELSKSFNQAAIIGQAGTNNNSAQLRQGGSKLLAVVAQEGS 76
 QY 98 RNNATIDOWNAKNSDITVGOYGGNNAALVNQATSDSSVMVROVGFNNATANQY 151
 DB 77 SNRAKIDQGDYNL-AYIDQAGSANDASISQAGYNTAMIIQKSGNKANITQY 129

RESULT 10
 C90806

minor curlin subunit precursor CsgB [imported] - Escherichia coli (strain O157:H7, subst
 C;Species: Escherichia coli
 C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
 C;Accession: C90806
 R;Hayaishi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
 S;Gawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 8, 11-22, 2001
 A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen
 A;Reference number: A99629; MUID:21156231; PMID:11258796
 A;Accession: C90806
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-151 <HAY>
 A;Cross-references: GB:BA000007; PIDN:BA34842.1; PID:G13360879; GSPDB:GN00154
 A;Experimental source: strain O157:H7, substrain RIMD 0509952
 C;Genetics:
 A;Gene: ECs1419

Query Match 12.5%; Score 96.5; DB 2; Length 151;
 Best Local Similarity 28.1%; Pred. No. 0.29; Mismatches 15; Conservative 32; Indels 5; Gaps 2;
 Matches 32; Conservative 15; Mismatches 62; Indels 5; Gaps 2;

QY 38 SSGPDTLSIYQGSANAALALQSDARKYDQVLVTRVTHEMAHAGQADNSTIELTQNGF 97
 DB 21 AAGYDLANSEYNF----AVNELSKSFNQAAIIGQAGTNNNSAQLRQGGSKLLAVVAQEGS 76
 QY 98 RNNATIDOWNAKNSDITVGOYGGNNAALVNQATSDSSVMVROVGFNNATANQY 151
 DB 77 SNRAKIDQGDYNL-AYIDQAGSANDASISQAGYNTAMIIQKSGNKANITQY 129

RESULT 11

G85565
 curlin minor chain precursor, CsgA homolog [imported] - Escherichia coli (strain O157:H7
 C;Species: Escherichia coli
 C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 C;Accession: G85565
 R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
 iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
 Nature 409, 529-533, 2001
 A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A;Reference number: A85480; MUID:21074935; PMID:11206551
 A;Accession: G85565
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-151 <STO>
 A;Cross-references: GB:AE005174; NID:G12514573; PIDN:AAGS5787.1; GSPDB:GN0145; UWGP:Z16
 A;Experimental source: strain O157:H7, substrain EDL933
 C;Genetics:
 A;Gene: csbB

Query Match 12.5%; Score 96.5; DB 2; Length 151;
 Best Local Similarity 28.1%; Pred. No. 0.29; Mismatches 15; Conservative 32; Indels 5; Gaps 2;
 Matches 32; Conservative 15; Mismatches 62; Indels 5; Gaps 2;

QY 38 SSGPDTLSIYQGSANAALALQSDARKYDQVLVTRVTHEMAHAGQADNSTIELTQNGF 97
 DB 21 AAGYDLANSEYNF----AVNELSKSFNQAAIIGQAGTNNNSAQLRQGGSKLLAVVAQEGS 76
 QY 98 RNNATIDOWNAKNSDITVGOYGGNNAALVNQATSDSSVMVROVGFNNATANQY 151
 DB 77 SNRAKIDQGDYNL-AYIDQAGSANDASISQAGYNTAMIIQKSGNKANITQY 129

RESULT 12

E95965
 hypothetical glycine-rich protein [imported] - Sinorhizobium meliloti (strain 1021) mag
 C;Species: Sinorhizobium meliloti
 C;Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
 C;Accession: E95965
 R;Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan
 Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
 A;Title: The complete sequence of the 1.683-kb pSymB megaplasmid from the N2-fixing endo

A;Reference number: A95842; MUID:21396508; PMID:11481431
A;Accession: E95965
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2174 <KUR>
A;Cross-references: GB:AL591985; PIDN:CA493389.1; PID:g15140875; GSPDB:GN00167
A;Experimental source: strain 1021, megaplasmid pSymb
R;Gailbert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler, P.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.; L.; Hyman, R.W.; Jones, T.
Science 293, 669-672, 2001
A;Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A;Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.
A;Reference number: A96039; MUID:21368234; PMID:11474104
A;Contents: annotation
C;Genetics:
A;Gene: Smb21548
A;Genome: plasmid

Query Match 12.4%; Score 96; DB 2; Length 2174;
Best Local Similarity 27.0%; Pred. No. 6.5;
Matches 40; Conservative 18; Mismatches 54; Indels 36; Gaps 7;

QY 11 AIVVSGSALAGVWPO--WGGGNGHNGGNSGPDSTLSIYQYGS-----ANAA----- 56
DB 693 AIAIAGAGAVILAQSIGGGG--GGNATGGAGGCGFQIGGGGGGGYANTANYGFK 749
QY 57 -LALQSDARKYDQVLRVVTTHMAHAGQAGDNSTIELTQNGFRNNATIDQWNAKNSDIT 115
DB 750 -GLTLTTQGSAAAGVAVQSV-----GGGGGTGGTASSYSAGIGFTAS-----VAV 793
QY 116 QYGGNNAA--LVNQTSADSSVMVRQV 141
DB 794 GGTGNGGAGGEVSVSLTDSAIRTQGG 821

RESULT 13
S11672
ice nucleation protein - Xanthomonas campestris
C;Species: Xanthomonas campestris
C;Date: 21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 26-Aug-1999
C;Accession: S11672
R;Zhao, J.; Orser, C.S.
Mol. Gen. Genet. 223, 163-166, 1990
A;Title: Conserved repetition in the ice nucleation gene *inx* from Xanthomonas campestris
A;Reference number: S11672; MUID:91080859; PMID:2259339
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1567 <ZHA>
A;Cross-references: EMBL:X52970; NID:g48531; PIDN:CAA37140.1; PID:g48532
C;Superfamily: ice nucleation protein

Query Match 12.0%; Score 92.5; DB 2; Length 1567;
Best Local Similarity 25.8%; Pred. No. 8.9;
Matches 46; Conservative 27; Mismatches 55; Indels 49; Gaps 11;

QY 14 VSGSALAG-----VVPQWGG--GGNHN-----GGGNSGPDSTLSIYQYGSANAALAL 59
DB 205 VYGSTLTADOSRLVAGVGTETAGDHSDDLGIAGVGTAGSDSI-LAGVGTQTAGR 263
QY 60 QSDARKYDQVLT-----RVVTHEMAHAGQAGDNSTIELTQNGFRNNATI----- 103
DB 264 SLTLAGYGTQTQAQGSRLTSGYGTATSGSDSAVI-----SGYGTQTAGSESSLTAGY 319
QY 104 -DOWNAKNSDITVQYGV-----GNNALV-----NOTASDSSVMVRQVGFNNATNQ 150
DB 320 STQTARKGSDITAG-YGSTAGSDSALIAGVGTQTAGSESSLT--AGVGTQTARK 374

RESULT 14
S07053

ice nucleation protein *inaA* - Erwinia ananas
C;Species: Erwinia ananas
C;Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 26-Aug-1999
C;Accession: S07053
R;Bebe, K.; Watabe, S.; Emori, Y.; Watanabe, M.; Arai, S.
FEBS Lett. 258, 297-300, 1989
A;Title: An ice nucleation active gene of *Erwinia ananas*. Sequence similarity to those of *Erwinia caroliniana* and *Erwinia ictioerythrae*.
A;Reference number: S07053; MUID:90092494; PMID:2590995
A;Accession: S07053
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-1322 <AB>
A;Cross-references: GB:X17316; NID:g296095; PIDN:CAA35194.1; PID:g296096
C;Superfamily: ice nucleation protein

Query Match 11.8%; Score 91.5; DB 2; Length 1322;
Best Local Similarity 29.1%; Pred. No. 8.9;
Matches 37; Conservative 19; Mismatches 22; Indels 49; Gaps 9;

QY 34 GCGNSGPDSTLSIYQYGSANAALQSDARKYDQVLRVVTTHMAHAGQAGDNSTIELT 93
DB 933 GSTSTAGPSSL-IAGYGTQTA-----GYSILT-----AGYS-----T 967
QY 94 QNGFRNNATIDQWNAKNSDITVQYGV-----GNNALV-----NOTASDSSVMVRQVGF 143
DB 968 QTG-----QENSLLTTG-YGSTAGVSSLIAGYGTQTATSPKSTLM--AGYG 1013
QY 144 NNATANO 150
DB 1014 SSQTARE 1020

RESULT 15

PL0221

leishmanolysin (EC 3.4.24.36) precursor [validated] - Leishmania major
N;Alternate names: promastigote surface proteinase; surface endopeptidase glycoprotein
C;Species: Leishmania major
C;Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 15-Sep-2000
C;Accession: PL0221; A27598; A60648
R;Button, L.L.; McMaster, W.R.
J. Exp. Med. 171, 589, 1990
A;Reference number: PL0221
A;Contents: erratum
A;Accession: PL0221
A;Molecule type: DNA
A;Residues: 1-602 <BUT>
A;Cross-references: GB:Y00647; NID:g9554; PIDN:CAA68673.1; PID:g9555
A;Note: this is a revision to the sequence from reference A27598
R;Button, L.L.; McMaster, W.R.
J. Exp. Med. 167, 724-729, 1988
A;Title: Molecular cloning of the major surface antigen of Leishmania.
A;Reference number: A27598; MUID:88154764; PMID:3346625
A;Accession: A27598

A;Status: significant sequence differences

A;Molecule type: DNA

R;Bouvier, J.; Bordier, C.; Vogel, H.; Reichelt, R.; Etges, R.

Mol. Biochem. Parasitol. 37, 235-246, 1989

A;Title: Characterization of the promastigote surface protease of Leishmania as a membranal protein.
A;Reference number: A60648; MUID:90114330; PMID:2608099
A;Accession: A60648

A;Molecule type: protein

A;Residues: 101, E', 103-118, SV', 121-123 <BOU>

A;Experimental source: strain LEM513

R;Schlagenhaut, E.; Etges, R.; Metcalf, P.

submitted to the Brookhaven Protein Data Bank, March 1997

A;Reference number: A68135; PDB:1LML

A;Contents: annotation; X-ray crystallography, 1.86 angstroms, residues 100-407, 412-498

A;Note: strain JRC-L119

C;Complex: homodimer

C;Function:

A;Description: catalyzes the hydrolysis of peptide bonds between two hydrophobic residues

A;Note: the activated form can activate the proenzyme form

C;Superfamily: leishmanolysin

C:Keywords: blocked carboxyl end; cell adhesion; glycoprotein; homodimer; hydrolase; lip
F:1-39/Domain: signal sequence #status predicted <SIG>
F:40-100/Domain: activation peptide #status predicted <ATP>
F:101-577/Product: leishmanolysin #status experimental <MAT>
F:578-602/Domain: carboxyl-terminal propeptide #status predicted <CTP>
F:148-264,268-334/Binding site: zinc, catalytic (Cys, His, His, His) (inhibited) #status
F:100-101/Cleavage site: Val-Val (autolytic) #status experimental
F:125-142,191-230,334-366,393-455,496-425,415-489,446-530,515-565,535-559/Disulfide bond
F:264,268,334/Binding site: zinc, catalytic (His) (active) #status experimental
F:265/Active site: Glu #status predicted
F:300,407/Binding site: carbohydrate (Asn) (covalent) #status experimental
F:577/Modified site: GPI-anchored ethanolamine amidated carboxyl end (Asn in mature form

```
Query Match      11.8%; Score 91; DB 1; Length 602;
Best Local Similarity 37.5%; Pred.No. 4;
Matches         27; Conservative    6; Mismatches   15; Indels   24; Gaps   3;
```

QY 63 ARKYDLVTRVVTHMAHA---GQGADNSTIELTONGFRNNATIDQWNAKSDITV---115
| : ||||| | | | | | : : | | : | | : |
Db 251 ASRYDVLTRVVTHMAHALGSGPFEDARIV-----ANPNVRNGKDFVPVINS 301

QY 116 -----GQYG 119
| | |
Db 302 STAVAKAREOYG 313

Search completed: August 2, 2004, 14:56:23
Job time : 9.4 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:36:12 ; Search time 5.3 seconds
(without alignments)

1483.508 Million cell updates/sec

Title: US-09-543-407-20

Perfect score: 774

Sequence: 1 MKLLKVAFAAIVVSGSLA.....DSSVMVRQVGFNNATANQV 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	689	89.0	151	1	CSGA_SALTY
2	528	68.2	151	1	CSGA_ECOLI
3	506.5	65.4	152	1	CSGA_ECO57
4	96.5	12.5	151	1	CSGB_ECOLI
5	92.5	12.0	1567	1	ICEN_XANCT
6	91.5	11.8	1322	1	ICEA_PANAN
7	91	11.8	602	1	GP63_LEIMA
8	88.5	11.4	151	1	CSGB_SALTY
9	88.5	11.4	151	1	GP63_SALTY
10	88	11.4	590	1	GP63_LEIDO
11	88	11.4	599	1	GP63_LEICH
12	88	11.4	646	1	GP63_LEIME
13	86.5	11.2	369	1	PST3_MYCAV
14	84.5	10.9	188	1	FMF2_ECOLI
15	84.5	10.9	1034	1	ICEN_PANAN
16	84.5	10.9	1258	1	ICEN_ERWHE
17	83	10.7	678	1	YF48_MYCTU
18	82.5	10.7	592	1	CSA_CITFR
19	82.5	10.7	641	1	IND_ARTGO
20	82.5	10.7	1028	1	OVO_DROME
21	81.5	10.5	392	1	HME1_HUMAN
22	81.5	10.5	1317	1	N445_YEAST
23	81	10.5	1656	1	OYEB_EICJA
24	80.5	10.4	1210	1	ICEN_PSEFL
25	78.5	10.1	472	1	SACB_BACAM
26	78.5	10.1	487	1	Y442_MYCTU
27	78	10.1	147	1	HPAA_CAUCR
28	78	10.1	363	1	PER_COPCI
29	77.5	10.0	254	1	PRIO_MESAU
30	77.5	10.0	364	1	PER_ARTPA
31	77.5	10.0	499	1	YRZ5_METTF
32	77	9.9	370	1	PST3_MYCTU
33	76	9.8	252	1	PRIO_RABIT

RESULT 1

CSGA_SALTY STANDARD; PRT; 151 AA.
ID AC P55225;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Major curlin subunit precursor (Fimbrin SEF17).
GN CSGA OR AGFA OR STM144 OR STM181 OR STM1776.
OS Salmonella typhimurium,
OS Salmonella typhi, and
OS Salmonella enteritidis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602, 601, 592;
[1]
SEQUENCE FROM N.A.
RP SPECIES=S.typhimurium; STRAIN=SR-11;
RC MEDLINE=98117058; PubMed=9457880;
RA Romling U., Bian Z., Hammar M., Sierralta W.D., Normark S.;
RT "Curli fibers are highly conserved between Salmonella typhimurium and
Escherichia coli with respect to operon structure and regulation.";
RL J. Bacteriol. 180:722-731(1998).
[2]
SEQUENCE FROM N.A.
RP SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
RC MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
Leonard S., Nguyen K., Scott K., Holmes A., Grewal N., Mulvaney E.,
Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
LT2.";
RL Nature 413:852-856(2001).
[3]
SEQUENCE FROM N.A.
RP SPECIES=S.typhi; STRAIN=CT18;
RC MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
Baker S., Basham D., Brooks K.M., Dowd L., White N., Farrar J.,
Cronin A., Davis P., Davies R., Chillingworth T., Connor P.,
Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
Whitehead S., Barrrell B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
[4]
SEQUENCE FROM N.A.
RP SPECIES=S.typhi; STRAIN=Ty2 / ATCC 700931;
RC MEDLINE=22531367; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
Burland V., Kodoyanni V., Schwartz D.C., Blattner F.R.;

ALIGNMENTS

34	76	9.8	504	1	FLIC_SALMC	Q06981 salmonella
35	76	9.8	955	1	TRU_DROME	Q8in81 drosophila
36	76	9.8	1115	1	TCB2_CHLRE	Q8vxp3 chlamydomon
37	75.5	9.8	254	1	PRIO_CRIGR	Q60506 cricetus
38	75.5	9.8	576	1	CEA7_ECOLI	Q47112 escherichia
39	75.5	9.8	1196	1	ICBV_PSEEX	Q33479 pseudomonas
40	75	9.7	172	1	CH18_DROME	P07184 drosophila
41	75	9.7	389	1	ACUC_STAAM	Q99tc9 staphylococ
42	74.5	9.6	163	1	HCV_NATPH	P39442 natronomona
43	74.5	9.6	370	1	PGLI_PENOL	Q9V834 penicillium
44	74.5	9.6	663	1	DUS8_MOUSE	O09112 mus musculus
45	74.5	9.6	760	1	YBIL_ECOLI	P73780 escherichia

RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
and CT18.";
RN J. Bacteriol. 185:2330-2337(2003).
[5]
RP SEQUENCE FROM N.A.
RC SPECIES=Salmonella enterica; STRAIN=27655-3B;
RX MEDLINE=96146512; PubMed=8550497;
RA Collinson S.K., Clouthier S.C., Doran J.L., Baner P.A., Kay W.W.;
RT "Salmonella enteritidis agfBAC operon encoding thin, aggregative
fimbriae.";
RN J. Bacteriol. 178:662-667(1996).
[6]
RP SEQUENCE OF 21-151 FROM N.A.
RC SPECIES=Salmonella enteritidis; STRAIN=27655-3B;
RX MEDLINE=94013373; PubMed=8104955;
RA Doran J.L., Collinson S.K., Burian J., Sarlos G., Todd E.C.D.,
RM Munro C.K., Kay W.W., Baner P.A., Peterkin P.I., Kay W.W.;
RT "DNA-based diagnostic tests for Salmonella species targeting agfA,
the structural gene for thin, aggregative fimbriae.";
RN J. Clin. Microbiol. 31:2263-2273(1993).
[7]
RP SEQUENCE OF 21-33.
RC SPECIES=Salmonella enteritidis; STRAIN=27655-3B;
RX MEDLINE=91310586; PubMed=1677357;
RA Collinson S.K., Emeody L., Mueller K.-M., Trust T.J., Kay W.W.;
RT "Purification and characterization of thin, aggregative fimbriae from
Salmonella enteritidis.";
RN J. Bacteriol. 173:4773-4781(1991).
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
FIBRONECTIN.
CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
DR EMBL; AF002301; CAA05317.1; -;
DR EMBL; AE008749; AAL20074.1; -;
DR EMBL; AL627269; CAD08268.1; -;
DR EMBL; AE016840; AAC09399.1; -;
DR EMBL; U43280; AAC43599.1; -;
DR FIC; JC6039; JC6039.
DR StyGene; SGI0608; csgA.
KW Signal; Complete proteome.
FT SIGNAL 1 20
FT CHAIN 21 151 MAJOR CURLIN SUBUNIT.
FT CONFLICT 134 151 SVMVRQVGFNNATANYQ -> DSYTVQVAS (IN
REF. 6)
SQ SEQUENCE 151 AA; 15305 MW; 87DADCD168621359 CRC64;
Query Match 89.0%; Score 689; DB 1; Length 151;
Best Local Similarity 90.7%; Pred. No. 3.9e-53;
Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVVSGSALAGVFPQWGGGNGHNGSSGPDSTLSIYQYCSANAALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVFPQWGGGNGHNGSSGPDSTLSIYQYCSANAALQ 60
QY 61 SDARKYDQLVTRVYTHENAHAQGGADNSTIELTQNGFNNATIDQWNAKNSDITVGYGG 120
DB 61 SDARKSETTITQSGYNGADVQGGADNSTIELTQNGFNNATIDQWNAKNSDITVGYGG 120
QY 121 NNAALVNTQADSSVMVRQVGFNNATANYQ 151
DB 121 NNAALVNTQADSSVMVRQVGFNNATANYQ 151

RESULT 2
CSGA_ECOLI
ID CSGA_ECOLI STANDARD; PRT; 151 AA.
AC P28307;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Major curlin subunit precursor.
GN CSGA OR B1042.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / W3110;
RX MEDLINE=932111294; PubMed=8459772;
RA Olsen A., Arngvist A.;
RT "The RpoS sigma factor relieves H-NS-mediated transcriptional
repression of csgA, the subunit gene of fibronectin-binding curli in
Escherichia coli.";
RN Mol. Microbiol. 7:523-536(1993).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MC4100;
RX MEDLINE=96414468; PubMed=8817489;
RA Hammar M., Arngvist A., Bian Z., Olsen A., Normark S.;
RT "Expression of two csg operons is required for production of
fibronectin- and Congo red-binding curli polymers in Escherichia coli
K-12.";
RN Mol. Microbiol. 18:661-670(1995).
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RN Science 277:1232-1244(1997).
[4]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Takemoto K., Inada T., Itoh T., Kajiura M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 12.7-28.0 min region on the linkage map.";
RN DNA Res. 3:137-155(1996).
[5]
RP SEQUENCE OF 21-40.
RC STRAIN=K12 / YMEI;
RX MEDLINE=93023873; PubMed=1357528;
RA Arngvist A., Olsen A., Pfeifer J., Russell D.G., Normark S.;
RT "The Crl protein activates cryptic genes for curli formation and
fibronectin binding in Escherichia coli HB101.";
RN Mol. Microbiol. 6:2443-2452(1992).
[6]
RP SEQUENCE OF 21-31.
RC STRAIN=K12 / YMEI;
RX MEDLINE=91310586; PubMed=1677357;
RA Collinson S.K., Emeody L., Trust T.J., Kay W.W.;
RT "Purification and characterization of thin, aggregative fimbriae from
Salmonella enteritidis.";
RN J. Bacteriol. 173:4773-4781(1991).
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
FIBRONECTIN.


```
CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L04979; AAA23616.1; -.
CC DR EMBL: X90754; CA62282.1; -.
CC DR EMBL: AE000205; AAC74126.1; -.
CC DR EMBL: D90741; BAA35832.1; -.
CC DR EMBL: D90742; BAA35840.1; -.
CC DR PIR: S70788; S70788.
CC ECoGene: EG11489; csGA.
CC FimBria; Signal; Complete proteome.
CC FT SIGNAL 1 20 MAJOR CURLIN SUBUNIT.
CC FT CHAIN 21 151 A -> E (IN REF. 1).
CC FT CONFLICT 7
CC SQ SEQUENCE 151 AA; 15049 MW; C003470D208D395F CRC64;
Query Match 68.2%; Score 528; DB 1; Length 151;
Best Local Similarity 68.9%; Pred. No. 3.6e-39;
Matches 104; Conservative 20; Mismatches 27; Indels 0; Gaps 0;
QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGDPSTLSIYQYGSANAALALQ 60
Db 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGDPSTLSIYQYGSANAALALQ 60
QY 61 SDARKYDOLVTRVVTHEMAHAGQGADNSTIELTQNGFRNNATIDOWNAKNSDITVGYGG 120
Db 61 TDARNSDITITQHGNGGADVQGGSDSSIDLTQRFNGSATLDQWNGKDSHMTVKQFG 120
QY 121 NNAALVNOTASDSSVMVQVFGNNATANQY 151
Db 121 GNGAAVDQTASNSVNVTVQVFGNNATAHQY 151
RESULT 3
CSGA_ECO57
ID CSGA ECO57 STANDARD; PRT; 152 AA.
AC Q93U24;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Major curlin subunit precursor.
GN CSGA OR Z1676 OR ECS1420.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / ATCC 43895;
RX MEDLINE=21218556; PubMed=11319125;
RA Ulrich G.A., Keen J.E., Elder R.O.;
RT "Mutations in the csGD promoter associated with variations in curli
RT expression in certain strains of Escherichia coli O157:H7."
RL Appl. Environ. Microbiol. 67:2367-2370(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
RT Nature 409:529-533(2001).
```

```
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii X., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12."
RL DNA Res. 8:11-22(2001).
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN.
CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF275733; AA53212.1; -.
CC DR EMBL: AE005315; ABG55788.1; -.
CC DR EMBL: AF002554; BAB34843.1; -.
CC DR PIR: D90806; D90806.
CC DR PIR: H85665; H85665.
CC FimBria; Signal; Complete proteome.
CC FT SIGNAL 1 20 BY SIMILARITY.
CC FT CHAIN 21 152 MAJOR CURLIN SUBUNIT.
CC SQ SEQUENCE 152 AA; 15099 MW; EE2D294DDE91243 CRC64;
Query Match 65.4%; Score 506.5; DB 1; Length 152;
Best Local Similarity 67.1%; Pred. No. 2.6e-37;
Matches 102; Conservative 21; Mismatches 28; Indels 1; Gaps 1;
QY 1 MKLLKVAAPAAIVVSGSALAGVVPQW-GGGGNGHNGGNSGDPSTLSIYQYGSANAALAL 59
Db 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGDPSTLSIYQYGSANAALAL 60
QY 60 QSDARKYDOLVTRVVTHEMAHAGQGADNSTIELTQNGFRNNATIDOWNAKNSDITVGYGG 119
Db 61 QADARNSDITITQHGNGGADVQGGSDSSIDLTQRFNGSATLDQWNGKDSHMTVKQFG 120
QY 120 GNGAAVDQTASNSVNVTVQVFGNNATANQY 151
Db 121 GNGAAVDQTASNSVNVTVQVFGNNATAHQY 152
RESULT 4
CSGB_ECOLI
ID CSGB_ECOLI STANDARD; PRT; 151 AA.
AC P39828;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Minor curlin subunit precursor.
GN CSGB OR B1041 OR Z1675 OR ECS1419.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MC4100;
RX MEDLINE=96414468; PubMed=8817489;
RA Hammar M., Arngvist A., Bian Z., Olsen A., Normark S.;
RT "Expression of two csG operons is required for production of
RT fibronectin- and congo red-binding curli polymers in Escherichia coli
```



```
SQ SEQUENCE 1567 AA; 152548 MW; C8B451D959ECAD63 CRC64;
Query Match 12.0%; Score 92.5; DB 1; Length 1567;
Best Local Similarity 25.8%; Pred. No. 3.1;
Matches 46; Conservative 27; Mismatches 56; Indels 49; Gaps 11;

QY 14 VSGSALAG-----VVPQGG---CGNHV-----GGNSSGPDSTLSIYQYGSNAALAL 59
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 205 VYGSTLTGADQSRVAGYGTETAGDHSDLIAGYGTGTAGSDSSI-LAGYGSTQTAAGR 263
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 60 QSDARKYDQVLT-----RVVTHENAHAGQAGDNNGTIELTONGFRNNATI----- 103
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 264 SFLTAGYGTGTAGSGRLTSGYGTATSGSDSAVI-----SGYGTGTAGSESSLTAGYG 319
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 104 -DOWNAKNSDITVQYQ-----GNNALV-----NOTASDSSVMVRQVGFNNATANQ 150
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 320 STQARKGSDITAG-YGSTGTAGSDSALIAGYGTGTAGSESSLT--AGYGSTQTARK 374
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 6
ID ICEA PANAN STANDARD; PRT; 1322 AA.
AC P20469;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ice nucleation protein Inaa.
GN INAA.
OS Pantoea ananas (Erwinia uredovora).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pantoea.
OX NCBI_TaxID=553;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90092494; PubMed=2599095;
RA Abe K., Watabe S., Emori Y., Watanabe M., Arai S.;
RT "An ice nucleation active gene of Erwinia ananas. Sequence similarity
RT to those of Pseudomonas species and regions required for ice
RT nucleation activity";
RL FEBS Lett. 258:297-300(1989).
CC -!- FUNCTION: Ice nucleation proteins enable bacteria to nucleate
CC crystallization in supercooled water.
CC -!- SUBCELLULAR LOCATION: Outer membrane (By similarity).
CC -!- DOMAIN: CONTAINS MANY IMPERFECT REPEATS OF THE CONSENSUS
CC OCTAPEPTIDE A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A
CC REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.
CC -!- SIMILARITY: Belongs to the bacterial ice nucleation protein
CC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X17316; CAA35194.1; .
CC DR PIR; S07053; S07053.
CC DR HSP; P06620; 1INA.
CC DR InterPro; IPR000258; Ice_nucleatn.
CC DR Pfam; PF00818; Ice_nucleation; 69.
CC DR PRINTS; PR00327; ICENUCLEATN.
CC DR PROSITE; PS00314; ICE_NUCLEATION; 49.
CC DR ICD; ICD00000; ICE_NUCLEATION; 49.
CC FT DOMAIN 162 1281 OCTAPEPTIDE PERIODICITY.
CC SEQUENCE 1322 AA; 131094 MW; 89B0EB24A837039 CRC64;

Query Match 11.8%; Score 91.5; DB 1; Length 1322;
Best Local Similarity 29.1%; Pred. No. 3.1;
Matches 37; Conservative 19; Mismatches 22; Indels 49; Gaps 9;

QY 34 GGNSSGPDSTLSIYQYGSNAALALQSDARKYDQVLTNRVTHENAHAGQAGDNSTIELT 93
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 933 GSTAGPDPSSL-IAGYGSTQTA-----CYNILT-----AGYGS-----T 967
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 94 QNGFRNNATIDOWNAKNSDITVQYQ-----GNNALV-----NOTASDSSVMVRQVGF 143
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 968 QTG-----QENSDLTTG-YGSTTAGYESSLIAGYGTGTATSPFKSTIM--AGYG 1013
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 144 NNATANQ 150
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1014 SSQTARE 1020
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 7
ID GP63 LEIMA STANDARD; PRT; 602 AA.
AC P08148; P15906;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Leishmanolysin precursor (EC 3.4.24.36) (Cell surface protease)
DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface
DE endopeptidase).
GN GP63.
OS Leishmania major.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A. AND SEQUENCE OF 101-123.
RX MEDLINE=88154764; PubMed=3346625;
RA Button L.L., McMaster W.R.;
RT "Molecular cloning of the major surface antigen of leishmania.";
RL J. Exp. Med. 167:724-729(1988).
RN [2]
RP REVISIONS.
RA Button L.L., McMaster W.R.;
RL J. Exp. Med. 171:589-589(1990).
RN [3]
RP GPI-ANCHOR.
RX MEDLINE=91009116; PubMed=2145267;
RA Schneider P., Ferguson M.A.J., McConville M.J., Mehler A.,
RA Homans S.W., Bordier C.;
RT "Structure of the glycosyl-phosphatidylinositol membrane anchor of
RT the Leishmania major promastigote surface protease.";
RL J. Biol. Chem. 265:16955-16964(1990).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).
RX MEDLINE=95406217; PubMed=7675788;
RA Schlagenhauf E., Etges R., Metcalf P.;
RT "Crystallization and preliminary X-ray diffraction studies of
RT Leishmanolysin, the major surface metalloproteinase from Leishmania
RT major.";
RL Proteins 22:58-66(1995).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.86 ANGSTROMS).
RX MEDLINE=98416698; PubMed=9739094;
RA Schlagenhauf E., Etges R., Metcalf P.;
RT "The crystal structure of the Leishmania major surface proteinase
RT Leishmanolysin."
RL Structure 6:1035-1046(1998).
CC -!- FUNCTION: Has an integral role during the infection of macrophages
CC in the mammalian host.
CC -!- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and
CC P1' and basic residues at P2 and P3'. A model nonapeptide is
CC cleaved at -Ala-Tyr-[Leu-Lys-Lys-.
CC -!- COFACTOR: Binds 1 zinc ion per subunit.
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -!- PMW: THE PHOSPHATIDYLINOSITOL MOIETY OF THE GPI-ANCHOR CONTAINS A
CC FULLY SATURATED, UNBRANCHED 1-O-ALKYL CHAIN (MAINLY C24:0) AND A
CC MIXTURE OF FULLY SATURATED UNBRANCHED 2-O-ACYL CHAINS (C12:0,
CC C14:0, C16:0, And C18:0).
CC -!- SIMILARITY: Belongs to peptidase family M8.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; Y00647; CAA68673.1; --
 DR PIR; PLO221; PLO221.
 DR PDB; 1LML; 17-SEP-97.
 DR MEROPS; M08.001; --
 DR InterPro; IPR006025; Pept_M_Zn_BS.
 DR InterPro; IPR001577; Peptidase_M8.
 DR Pfam; PF01457; Peptidase_M8; 1.
 DR PRINTS; PR00782; LSHMANOLYSIN.
 DR PROSITE; PS00142; ZINC_PROTEASE; 1.
 KW Hydrolase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;
 KW Zymogen; Signal; Cell adhesion; GPI-anchor; 3D-structure; Lipoprotein.
 FT SIGNAL 1 39 POTENTIAL.
 FT PROPEP 40 100 ACTIVATION PEPTIDE.
 FT CHAIN 101 577 LEISHMANOLYSIN.
 FT PROPEP 578 602 REMOVED IN MATURE FORM.
 FT METAL 264 264 ZINC (CATALYTIC).
 FT ACT_SITE 265 265
 FT METAL 268 268 ZINC (CATALYTIC).
 FT METAL 334 334 ZINC (CATALYTIC).
 FT DISULFID 125 142
 FT DISULFID 191 230
 FT DISULFID 314 386
 FT DISULFID 393 455
 FT DISULFID 406 425
 FT DISULFID 415 489
 FT DISULFID 466 510
 FT DISULFID 515 565
 FT DISULFID 535 558
 FT CARBOHYD 300 300
 FT CARBOHYD 407 407
 FT LIPID 577 577
 FT STRAND 101 102
 FT STRAND 107 108
 FT STRAND 111 114
 FT HELIX 116 119
 FT TURN 121 122
 FT TURN 128 129
 FT STRAND 131 133
 FT STRAND 139 141
 FT HELIX 144 146
 FT HELIX 150 158
 FT TURN 159 159
 FT HELIX 160 169
 FT TURN 170 171
 FT STRAND 172 174
 FT STRAND 177 178
 FT STRAND 180 181
 FT TURN 189 190
 FT HELIX 191 193
 FT HELIX 198 202
 FT TURN 203 203
 FT STRAND 205 206
 FT STRAND 210 215
 FT TURN 221 222
 FT STRAND 226 232
 FT TURN 234 235
 FT STRAND 238 244
 FT STRAND 247 249
 FT HELIX 256 269
 FT TURN 270 271
 FT HELIX 274 279
 FT TURN 280 281
 FT STRAND 283 286
 FT HELIX 289 291
 FT STRAND 296 299
 FT HELIX 302 312

FT TURN 313 313
 FT TURN 316 316
 FT STRAND 320 322
 FT TURN 328 328
 FT STRAND 335 335
 FT TURN 337 339
 FT TURN 341 342
 FT STRAND 343 343
 FT TURN 344 345
 FT STRAND 353 353
 FT HELIX 356 364
 FT TURN 365 366
 FT STRAND 369 370
 FT HELIX 372 374
 FT TURN 380 383
 FT HELIX 386 390
 FT STRAND 394 395
 FT TURN 396 397
 FT STRAND 398 399
 FT TURN 402 404
 FT STRAND 413 414
 FT TURN 417 418
 FT STRAND 421 425
 FT STRAND 428 429
 FT HELIX 435 437
 FT TURN 443 444
 FT STRAND 445 446
 FT TURN 450 454
 FT STRAND 458 465
 FT TURN 466 467
 FT HELIX 470 472
 FT TURN 475 477
 FT HELIX 478 480
 FT TURN 485 486
 FT STRAND 487 494
 FT STRAND 496 496
 FT STRAND 506 516
 FT TURN 517 520
 FT STRAND 521 525
 FT TURN 527 528
 FT STRAND 533 534
 FT TURN 537 538
 FT STRAND 540 542
 FT HELIX 543 545
 FT TURN 546 546
 FT STRAND 550 550
 FT TURN 552 553
 FT STRAND 555 557
 FT HELIX 561 565
 FT TURN 566 567
 FT HELIX 569 572
 FT TURN 573 573
 SQ SEQUENCE 602 AA; 63953 MW; 982EF3245D87C43E CRC64;

Query Match 11.8%; Score 91; DB 1; Length 602;

Best Local Similarity 37.5%; Pred. No. 1.4;

Matches 27; Conservative 6; Mismatches 15; Indels 24; Gaps 3;

QY 63 ARKYDQVTRVVTHEMAHA-----GQADNSTIELTQNGFRNNATIDQWNAKSDITV--- 115
 Db 251 ASRYDQVTRVVTHEMAHAALGSGPFEDARIV-----ANPNVRGKNFVDFVINS 301

QY 116 -----GQYG 119

Db 302 STAVAKAREQYG 313

RESULT 8

CSGB_SALTI ID_CSGB_SALTI STANDARD; PRT; 151 AA.

AC Q827M3;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

```

DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Minor curlin subunit precursor.
CSGB OR STV1180 OR T1777.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
ON NCBI_TaxID=601;
RX [1]
RN SEQUENCE FROM N.A.
RC STRAIN=CT18;
RP MEDLINE=21534947; PubMed=11677608;
RX Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K.M., Chillingworth T., Connerton P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell I., Hamlin N., Haque A., Hien T.F., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrett B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
[2]
RN SEQUENCE FROM N.A.
RP STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Xodoyanni V., Schwartz D.C., Blattner F.R.;
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RT and CT18.";
RL J. Bacteriol. 185:2330-2337(2003).
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
CC CURLIN MONOMERS.
CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; AL627269; CAD08267.1; -
DR DR EMBL; AB016940; AB069400.1; -
KW Fibria; Signal; Complete proteome.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 151 MINOR CURLIN SUBUNIT.
SQ SEQUENCE 151 AA; 16254 MW; 161C54326E573495 CRC64;
-----
Query Match 11.4%; Score 88.5; DB 1; Length 151;
Best Local Similarity 31.0%; Pred No. 0.48;
Matches 26; Conservative 13; Mismatches 42; Indels 3; Gaps 2;

QY 69 LVTRVVTHEMAHAGQGADNSTIELTQTGFRRNATIDQWNAKSDIT-VGQYGNNAAVN 127
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 48 IIQGVGTDSARVRQEGLSKLSVISQEGENNRKAVDQ--AGNYNFAYIEQTGNANDASIS 105
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 128 QTASDSSVMVROVFNGNNATANQY 151
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB .106 QSAYGNSAIIQKGSGNKANITQY 129
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 9
CSGB_SALTY STANDARD; PRT; 151 AA.
AC PS5226;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)

```

```

RESULT 10
GP63_LEIDO STANDARD; PRT; 590 AA.
ID P23223;
AC 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DE Leishmanolysin precursor (EC 3.4.24.36) (Cell surface protease)
DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface
DE endopeptidase).
GN GP63.
OS Leishmania donovani.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5661;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LV9;
RX MEDLINE=92107220; PubMed=1762629;
RA Webb J.R., Button L.L., McMaster R.W.;
RT "Heterogeneity of the genes encoding the major surface glycoprotein
RT of Leishmania donovani."
RL Mol. Biochem. Parasitol. 48:173-184 (1991).
CC -!- FUNCTION: Has an integral role during the infection of macrophages
CC in the mammalian host.
CC -!- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and
CC P1', and basic residues at P2 and P3'. A model nonapeptide is
CC cleaved at -Ala-Tyr-|-Leu-Lys-Lys-.
CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -!- SIMILARITY: Belongs to peptidase family M8.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M60048; AAA29244.1; -.
CC DR HSSP; P08148; 1LML.
CC DR MEROPS; M08.001; -.
CC DR InterPro; IPR006025; Pept_M_Zn_BS.
CC DR InterPro; IPR001577; Peptidase_M8.
CC DR Pfam; PF01457; Peptidase_M8; 1.
CC DR PRINTS; PR00782; LSHMANOLYSIN.
CC DR PROSITE; PS00142; ZINC_PROTEASE; 1.
CC KW Hydrolase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;
CC Zymogen; Signal; Cell adhesion; GPI-anchor; Lipoprotein.
CC SIGNAL 1 39 POTENTIAL.
CC FT PROPEP 40 87 ACTIVATION PEPTIDE.
CC FT CHAIN 88 565 LEISHMANOLYSIN.
CC FT PROPEP 566 590 REMOVED IN MATURE FORM (BY SIMILARITY).
CC FT METAL 251 251 ZINC (CATALYTIC) (BY SIMILARITY).
CC FT ACT_SITE 252 252 ZINC (CATALYTIC) (BY SIMILARITY).
CC FT METAL 255 255 ZINC (CATALYTIC) (BY SIMILARITY).
CC FT METAL 321 321 ZINC (CATALYTIC) (BY SIMILARITY).
CC FT DISULFID 112 129 BY SIMILARITY.
CC FT DISULFID 178 217 BY SIMILARITY.
CC FT DISULFID 301 373 BY SIMILARITY.
CC FT DISULFID 380 443 BY SIMILARITY.
CC FT DISULFID 393 412 BY SIMILARITY.
CC FT DISULFID 402 477 BY SIMILARITY.
CC FT DISULFID 454 498 BY SIMILARITY.
CC FT DISULFID 503 553 BY SIMILARITY.
CC FT DISULFID 523 546 BY SIMILARITY.
CC FT CARBOHYD 287 287 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT LIPID 565 565 GPI-anchor amidated asparagine (By
CC similarity).
CC SEQUENCE 590 AA; 62950 MW; 0FB315D29659F58 CRC64;

Matches 17; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 63 ARKYDQLVTRVVTHEVAHA 81
Db 238 ASRYDQLVTRVVTHEVAHA 256
| : ||||| ||||| |||||
| : ||||| ||||| |||||

RESULT 11
GP63_LEICH STANDARD; PRT; 599 AA.
ID GP63_LEICH
AC P15706;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Leishmanolysin precursor (EC 3.4.24.36) (Cell surface protease)
DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface
DE endopeptidase).
GN GP63.
OS Leishmania chagasi.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=44271;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90205976; PubMed=2320059;
RA Miller R.A., Reed S.G., Parsons M.;
RT "Leishmania gp63 molecule implicated in cellular adhesion lacks an
RT Arg-Gly-Asp sequence."
RL Mol. Biochem. Parasitol. 39:267-274 (1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92112918; PubMed=1370484;
RA Ramamoorthy R., Donelson J.E., Paetz K.E., Maybodi M., Roberts S.C.,
RA Wilson M.E.;
RT "Three distinct RNAs for the surface protease gp63 are differentially
RT expressed during development of Leishmania donovani chagasi
RT promastigotes to an infectious form."
RL J. Biol. Chem. 267:1888-1895 (1992).
CC -!- FUNCTION: Has an integral role during the infection of macrophages
CC in the mammalian host.
CC -!- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and
CC P1', and basic residues at P2 and P3'. A model nonapeptide is
CC cleaved at -Ala-Tyr-|-Leu-Lys-Lys-.
CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -!- SIMILARITY: Belongs to peptidase family M8.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M80672; AAA29238.1; -.
CC DR EMBL; M28527; AAA29235.1; -.
CC DR PIR; A44951; A44951.
CC DR HSSP; P08148; 1LML.
CC DR MEROPS; M08.001; -.
CC DR InterPro; IPR006025; Pept_M_Zn_BS.
CC DR InterPro; IPR001577; Peptidase_M8.
CC DR Pfam; PF01457; Peptidase_M8; 1.
CC DR PRINTS; PR00782; LSHMANOLYSIN.
CC DR PROSITE; PS00142; ZINC_PROTEASE; 1.
CC KW Hydrolase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;
CC Zymogen; Signal; Cell adhesion; GPI-anchor; Lipoprotein.
CC SIGNAL 1 39 POTENTIAL.
CC FT PROPEP 40 97 ACTIVATION PEPTIDE.
CC FT CHAIN 98 574 LEISHMANOLYSIN.
CC FT PROPEP 575 599 REMOVED IN MATURE FORM (BY SIMILARITY).
CC FT METAL 261 261 ZINC (CATALYTIC) (BY SIMILARITY).
CC FT ACT_SITE 262 262 BY SIMILARITY.
CC FT METAL 265 265 ZINC (CATALYTIC) (BY SIMILARITY).

```

Query Match 11.4%; Score 88; DB 1; Length 590;
Best Local Similarity 89.5%; Pred. No. 2.5;

```
FT METAL 331 ZINC (CATALYTIC) (BY SIMILARITY).
FT DISULFID 122 BY SIMILARITY.
FT DISULFID 188 BY SIMILARITY.
FT DISULFID 311 BY SIMILARITY.
FT DISULFID 390 BY SIMILARITY.
FT DISULFID 403 BY SIMILARITY.
FT DISULFID 412 BY SIMILARITY.
FT DISULFID 463 BY SIMILARITY.
FT DISULFID 512 BY SIMILARITY.
FT DISULFID 532 BY SIMILARITY.
FT CARBOHYD 297 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 394 N-LINKED (GLCNAC. .) (POTENTIAL).
FT LIPID 574 GPI-anchor amidated asparagine (By similarity).
SQ SEQUENCE 599 AA; 63048 MW; 746730AE8E2A2E7C CRC64;

Query Match 11.4%; Score 88; DB 1; Length 599;
Best Local Similarity 89.5%; Pred. No. 2.6;
Matches 17; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 63 ARKYDQLVTRVVTHEMAHA 81
DB 248 ASRYDQLVTRVVTHEMAHA 266

RESULT 12
GP63 LEIME STANDARD; PRT; 646 AA.
ID GP63 LEIME STANDARD; PRT; 646 AA.
AC P43150;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Leishmanolysin C1 precursor (SC 3.4.24.36) (Cell surface protease)
DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface endopeptidase).
DE GP63-C1.
GN GP63-C1.
OS Leishmania mexicana.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OC NCBI_TaxID=5665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MNYC/BZ/62/M379;
RX MEDLINE=93149206; PubMed=8426614;
RA Medina-Acosta E., Kress R.E., Russell D.G.;
RT "Structurally distinct genes for the surface protease of Leishmania mexicana are developmentally regulated."
RL Mol. Biochem. Parasitol. 57:31-46(1993).
CC -!- FUNCTION: Has an integral role during the infection of macrophages in the mammalian host.
CC -!- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and P1' and basic residues at P2 and P3'. A model nonapeptide is cleaved at -Ala-Tyr-|-Leu-Lys-Lys-.
CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC -!- DEVELOPMENTAL STAGE: Expressed in both the promastigote and the amastigote forms.
CC -!- SIMILARITY: Belongs to peptidase family M8.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC
CC EMBL; X64394; CAA45733.1; -.
CC DR PIR; S19916; S19916.
CC DR HSP; P08148; 1LML.
CC DR MEROPS; M08.001; -.
CC DR GlycoSuiteDB; P43150; -.
CC DR InterPro; IPR006025; Pept_M_Zn_BS.
CC DR InterPro; IPR001577; Peptidase M8.
CC Pfam; PF01457; Peptidase M8; 1.
```

```
DR PRINTS; PR00782; LSHMANOLYSIN.
DR PROSITE; PS00142; ZINC PROTEASE; 1.
KW Hydrolase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;
KW Zymogen; Signal; Cell adhesion; Multigene family.
FT SIGNAL 1 39 POTENTIAL.
FT PROPEP 40 102 ACTIVATION PEPTIDE (POTENTIAL).
FT CHAIN 103 646 LEISHMANOLYSIN C1.
FT METAL 266 266 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 267 267 BY SIMILARITY.
FT METAL 270 270 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 336 336 ZINC (CATALYTIC) (BY SIMILARITY).
FT DISULFID 127 144 BY SIMILARITY.
FT DISULFID 193 232 BY SIMILARITY.
FT DISULFID 316 388 BY SIMILARITY.
FT DISULFID 395 458 BY SIMILARITY.
FT DISULFID 408 427 BY SIMILARITY.
FT DISULFID 417 492 BY SIMILARITY.
FT DISULFID 469 513 BY SIMILARITY.
FT DISULFID 518 568 BY SIMILARITY.
FT DISULFID 538 561 BY SIMILARITY.
FT CARBOHYD 86 86 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 297 297 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 399 399 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 403 403 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 433 433 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 445 445 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 466 466 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 501 501 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 646 AA; 69054 MW; FE448DDC78C10B0A CRC64;

Query Match 11.4%; Score 88; DB 1; Length 646;
Best Local Similarity 89.5%; Pred. No. 2.8;
Matches 17; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 63 ARKYDQLVTRVVTHEMAHA 81
DB 253 ASRYDQLVTRVVTHEMAHA 271

RESULT 13
PST3 MYCAV STANDARD; PRT; 369 AA.
ID PST3 MYCAV STANDARD; PRT; 369 AA.
AC QSKK59;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Phosphate-binding protein 3 precursor (PBP-3) (PstS-3).
GN PST3.
OS Mycobacterium avium.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OC NCBI_TaxID=1764;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=969A45;
RA Carroll J.D., Wallace R.C., Keane J., Arbeit R.D.;
RT "Identification of Mycobacterium avium DNA sequences that encode exported proteins by using phoA gene fusions."
RT Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Required for binding-protein-mediated phosphate transport (By similarity).
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor (By similarity).
CC -!- SIMILARITY: Belongs to the pstS family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC
```

DR EMBL; AF137360; AAF74819.1; -;
 DR HSP; P06128; 1A54.
 DR InterPro; IPR000437; Prok lipoprot s.
 DR InterPro; IPR006059; SBP_bac_1.
 DR Pfam; PF01547; SBP_bac_1.
 DR PROSITE; PS00013; PROKAP_LIPOPROTEIN; 1.
 KW Phosphate transport; Transport; Membrane; Lipoprotein; Signal;
 KW Palmitate.
 FT SIGNAL 1 22 POTENTIAL.
 FT CHAIN 23 369 PHOSPHATE-BINDING PROTEIN 3.
 FT LIPID 23 23 N-palmitoyl cysteine (Potential).
 FT LIPID 23 23 S-disulphide-linked cysteine (Potential).
 SQ SEQUENCE 369 AA; 37225 MW; CB0EA0AC10F483EC CRC64;
 Query Match 11.2%; Score 86.5; DB 1; Length 369;
 Best Local Similarity 27.8%; Pred. No. 2;
 Matches 35; Conservative 21; Mismatches 49; Indels 21; Gaps 6;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQGSANAALALQ 60
 Db 1 MKLNFGAVLSVLSAGALV-----SGGSDNNGAGAGAGSSSKVSCGKXKALRASG 54
 QY 61 SDARKYDQLVTRVW-THEMAHAGQADNSTIELTQGFNNATIDQWNAKNSDITVGYG 119
 Db 55 STAQA--NMTFRVNAFEACPGQ-----TLNYTANG--SGAGISEFNGKQTD----FG 100
 QY 120 GNNAL 125
 Db 101 GSDSPL 106
 RESULT 14
 FMF2_ECOLI STANDARD; PRT; 188 AA.
 ID FMF2_ECOLI STANDARD; PRT; 188 AA.
 AC P02972;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE F7-2 fimbrial protein precursor (F7-2 pilin).
 GN F7-2 OR PAPA OR C3592.
 OS Escherichia coli, and
 OS Escherichia coli O6.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562, 217992;
 RN [1]_TaxID=562, 217992;
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85155489; PubMed=6152241;
 RA van Die I., Bergmans H.;
 RT "Nucleotide sequence of the gene encoding the F72 fimbrial subunit of
 RT a uropathogenic Escherichia coli strain.";
 RL Gene 32:83-90(1984).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92040048; PubMed=1682251;
 RA Denich K., Blyn L.B., Craiu A., Braaten B.A., Hardy J., Low D.A.,
 RA O'Hanley P.D.;
 RT "DNA sequences of three pAPA genes from uropathogenic Escherichia
 RT coli strains: evidence of structural and serological conservation.";
 RL Infect. Immun. 59:3849-3858(1991).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O6:Hi / CFT073 / ATCC 700928;
 RX MEDLINE=22388234; PubMed=12471157;
 RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
 RA Rasko D., Buckles E.I., Liou S.-R., Boutin A., Hackett J., Stroud D.,
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
 RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
 RT "Extensive mosaic structure revealed by the complete genome sequence
 RT of uropathogenic Escherichia coli.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
 CC -!- FUNCTION: Fimbriae (also called pili), polar filaments radiating
 CC from the surface of the bacterium to a length of 0.5-1.5

CC micrometers and numbering 100-300 per cell, enable bacteria to
 CC colonize the epithelium of specific host organs.
 CC -!- SIMILARITY: BELONGS TO THE FIMB/PAPA FAMILY OF FIMBRIA PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>;
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M12861; AAA23778.1; -;
 DR EMBL; M68060; AAA24278.1; -;
 DR EMBL; AE016766; AAN82040.1; ALT_INIT.
 DR PIR; A03496; YQECF2.
 DR InterPro; IPR008966; Adhes_bact.
 DR InterPro; IPR000259; Fimbrial.
 DR Pfam; PF00419; Fimbrial; 1.
 KW Fimbria; Signal; Complete proteome.
 FT SIGNAL 1 21
 FT CHAIN 22 188 F7-2 FIMBRIAL PROTEIN.
 FT DISULFID 43 82 PROBABLE.
 SQ SEQUENCE 188 AA; 19184 MW; 0EEF750CFD843157 CRC64;
 Query Match 10.9%; Score 84.5; DB 1; Length 188;
 Best Local Similarity 25.0%; Pred. No. 1.4; Indels 33; Gaps 8;
 Matches 45; Conservative 26; Mismatches 76;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNG-----GNSSGDPSTLSIYQYGS 52
 Db 2 IKSVIAGAVAMAVSPGAYAPTIPQGGKVTFTNGTVVDPAGIDAGADQSIDFGQV-- 59
 QY 53 ANAALALQSDA----RKYD-QLVTRVTVTH-ENAHAGQADNSTIELTQGFNNATIDQW 106
 Db 60 --SKLFENDGESQPKSFDIKINCIDITNFKAAGGGGARTGTSLTSGVSGSQSMDL 117
 QY 107 ---NAKNSDITVGYGGNNAALVNQTSADSSVM-----VROVGFNNATANQY 151
 Db 118 QTVGAINTAIVTDPHGKVKFDGATATGVSILVDGNTIHTFAAVRKGDSGNPVTGAF 177
 RESULT 15
 ICEN PANAN STANDARD; PRT; 1034 AA.
 ID ICEN PANAN STANDARD; PRT; 1034 AA.
 AC Q47879;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 18-OCT-2001 (Rel. 40, Last annotation update)
 DE Ice nucleation protein inau.
 GN INAU.
 OS Pantoea ananas (Erwinia uredovora).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Pantoea.
 OX NCBI_TaxID=553;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KUIN-3;
 RX MEDLINE=94264407; PubMed=7764866;
 RA Michigami Y., Watabe S., Abe K., Obata H., Arai S.;
 RT "Cloning and sequencing of an ice nucleation active gene of Erwinia
 RT uredovora";
 RL Biosci. Biotechnol. Biochem. 58:762-764(1994).
 CC -!- FUNCTION: Ice nucleation proteins enable bacteria to nucleate
 CC crystallization in supercooled water.
 CC -!- SUBCELLULAR LOCATION: Outer membrane.
 CC -!- DOMAIN: CONTAINS IMPERFECT REPEATS OF A CONSENSUS OCTAPEPTIDE
 CC A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A REGIONAL 48-RESIDUE
 CC PERIODICITY IS SUPERIMPOSED.
 CC -!- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE
 CC NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.
 CC -!- SIMILARITY: Belongs to the bacterial ice nucleation protein
 CC family.


```
CC ----- This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D14992; BAA03636.1; -
DR PIR; JC2143; JC2143.
DR HSRP; P06620; IINA.
DR InterPro; IPR000258; Ice_nucleatn.
DR Pfam; PF00818; Ice_nucleation; 51.
DR PRINTS; PR00327; ICENUCLEATN.
DR PROSITE; PS00314; ICE_NUCLEATION; 34.
KW Ice nucleation; Repeat; Outer membrane.
FT DOMAIN 162..993 OCTAPEPTIDE PERIODICITY.
SQ SEQUENCE 1034 AA; 103378 MW; FA222523D333EADD CRC64;

Query Match          10.9%; Score 84.5; DB 1; Length 1034;
Best Local Similarity 26.9%; Pred. NO. 9.6;
Matches 45; Conservative 17; Mismatches 66; Indels 39; Gaps 9;

Qy 12 IVVSGSALAGVVPQW--GGGNNHGGNSS-----GPDSTLSIYQYGSNAAL 57
Db 161 IATYGSTLSGTHQSLIAGYGSTETAGDSSTLIAGYGSTGTAGSDSTL-VAGYGSTQTAG 219

Qy 58 ALQSDARKYDQLVTRV----VTHEMAHAGQADNSTIETLQNGFRNNATID----- 104
Db 220 EESSQAGYGSTQTGMKGSDLTAGYGSTGTAGDDSSL-IAGYGSTQTAGDSSSLTAGYGS 278

Qy 105 -QWNAKNSDITVGYGGNNAALVNQTASDSSVMYRVQVGFNNATANQ 150
Db 279 TQTAQKGSDLTAG-YGSTGTA-----GADSSLI---AGYGSTQTAGE 316
```

Search completed: August 2, 2004, 14:49:30
Job time : 6.3 secs

mis Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:39:17 ; Search time 29.7 Seconds
(without alignments)
1604.150 Million cell updates/sec

Title: US-09-543-407-20
Perfect score: 774
Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVQVGFNNATANYQ 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues
Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : SPTREMBL_25:*
- 1: sp_archaea:*
 - 2: sp_bacteria:*
 - 3: sp_fungi:*
 - 4: sp_human:*
 - 5: sp_invertebrate:*
 - 6: sp_mammal:*
 - 7: sp_mac:*
 - 8: sp_organelle:*
 - 9: sp_phase:*
 - 10: sp_plant:*
 - 11: sp_rodent:*
 - 12: sp_virus:*
 - 13: sp_vertebrate:*
 - 14: sp_unclassified:*
 - 15: sp_rvirus:*
 - 16: sp_bacteriap:*
 - 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	680	87.9	152	2	Q33802	O33802 salmonella
2	582.5	75.3	150	2	Q7X243	Q7X243 citrobacter
3	552	71.3	149	2	Q7X240	Q7X240 citrobacter
4	503.5	65.1	152	16	Q8CW63	Q8CW63 escherichia
5	430.5	55.6	150	2	Q7X237	Q7X237 enterobacte
6	301	38.9	76	2	Q34069	Q34069 salmonella
7	122	15.8	29	2	Q9S3U5	Q9S3U5 escherichia
8	115	14.9	130	16	Q89U14	Q89U14 bradyrhizob
9	110	14.2	139	16	Q8EIH3	Q8EIH3 shewanella
10	106.5	13.8	502	16	Q8EIH4	Q8EIH4 shewanella
11	105.5	13.6	145	16	Q8UGN9	Q8UGN9 agrobacteri
12	105.5	13.6	153	16	Q89U16	Q89U16 bradyrhizob
13	99	12.8	573	10	Q9SAR2	Q9SAR2 arabidopsis
14	98.5	12.7	154	16	Q89U15	Q89U15 bradyrhizob
15	97.5	12.6	1765	16	Q7V8S5	Q7V8S5 prochloroco
16	97	12.5	157	16	Q88HG0	Q88HG0 pseudomonas

17 96.5 12.5 151 16 Q7UCZ1
18 96.5 12.5 160 16 Q8CW64
19 96.5 12.5 160 16 Q83RU7
20 96 12.4 2174 16 Q92U08
21 95 12.3 1422 16 Q8EFU3
22 94.5 12.2 438 16 Q82M56
23 94.5 12.2 598 5 Q25275
24 94.5 12.2 1209 16 Q89CK5
25 93 12.0 480 16 Q89EV2
26 92.5 12.0 91 2 Q8S3U8
27 92.5 12.0 151 2 Q7X244
28 92 11.9 644 5 Q43994
29 92 11.9 1408 16 Q8E833
30 91 11.8 645 16 Q7U1C5
31 91 11.8 646 16 Q53818
32 91 11.8 1410 16 Q8CM00
33 91 11.8 2734 16 Q89C73
34 90.5 11.7 329 3 Q8N1Z1
35 90 11.6 152 2 Q7X241
36 90 11.6 639 5 Q25274
37 90 11.6 644 5 Q8VNY9
38 90 11.6 714 16 Q7U5X6
39 90 11.6 1460 16 Q8K6B1
40 90 11.6 3056 16 Q7USQ0
41 90 11.6 3501 16 Q8Y106
42 90 11.6 3552 16 Q8XSD6
43 89.5 11.6 151 2 Q7X238
44 89.5 11.6 297 2 Q9Z1Y5
45 89.5 11.6 333 5 Q8T986

ALIGNMENTS

RESULT 1
Q33802 ID Q33802 PRELIMINARY; PRT; 152 AA.
AC Q33802;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE AgfA protein (Fragment).
GN AgfA.
OS Salmonella typhimurium.
CC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
CC Enterobacteriaceae; Salmonella.
CX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98053981; PubMed=9393832;
RA Sukupolvi S.S., Lorentz R.G., Gordon J.I., Bian Z., Pfeifer J.D.,
RA Normark S.J., Rhen M.;
RT "Expression of thin, aggregative fimbriae promotes interaction of
RT Salmonella typhimurium SR-11 with mouse small intestinal epithelial
RT cells.";
RL Infect. Immun. 65:5320-5325 (1997).
DR EMBL; AJ000514; CAA04151.1; -;
FT NON_TER
FT 152
SQ SEQUENCE 152 AA; 15401 MW; 9DA7DADC2364B006 CRC64;

Query Match 87.9%; Score 680; DB 2; Length 152;
Best Local Similarity 89.4%; Pred. No. 2.3e-48;
Matches 135; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

Qy 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNSGPDSTLSIYQYGSANAALAIQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNSGPDSTLSIYQYGSANAALAIQ 60
Qy 61 SDARKYDQVLRVTVTHEMAHAGGADNSTIETQNGFRNNATIDOWNAKNSDITVGYGG 120
Db 61 SDARKSETTITQSGYNGADVCGGADNSTIETQNGFRNNATIDOWNAKNSDITVGYGG 120
Qy 121 NNAALVNTQASDSSVMVQVGFNNATANYQ 151

```

Db      121 NNAALVNQTASDSSVMVRQVGFNNAPANQY 151
|||||
RESULT 2
Q7X243
ID      Q7X243      PRELIMINARY;      PRT;      150 AA.
AC      Q7X243;
DT      01-OCT-2003 (TReMBLrel. 25, Created)
DT      01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DT      01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE      Curlin-csgA protein.
GN      CSGA.
OS      Citrobacter sp. Fec2.
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC      Enterobacteriaceae; Citrobacter.
OX      NCBI_TaxID=213763;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=Fec2;
RA      Zogaj X., Bokranz W., Nimitz M., Romling U.;
RT      "Production of Cellulose and Curli Fimbriae by Members of the Family
RT      Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL      Infect. Immun. 72:4151-4158 (2003).
DR      EMBL; AJ515700; CAD56672.1; -.
SQ      SEQUENCE 150 AA; 15016 MW; 1D7141B8D6973DC6 CRC64;

Query Match      75.3%; Score 582.5; DB 2; Length 150;
Best Local Similarity 78.8%; Pred. No. 2.3e-40;
Matches 119; Conservative 10; Mismatches 21; Indels 1; Gaps 1;

QY      1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGSSGGPDSTLSIYQVGSANAALQ 60
|||||
Db      1 MKLLQVAAAPAAIVVSGSALAGVVPQWGGG- GGGSSSGPESTLSIYQVGSANAALQ 59
|||||

QY      61 SDARKYDQLVTRVVTHEMAHAGGADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGG 120
|||||
Db      60 SDARKSDTTIHONGFGNGADVQGGSDNSTIDLTQNGFKNNATIDQWNGKNSDITVQYGG 119
|||||

QY      121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
|||||
Db      120 HNAALVNQTASDSSVLVHQVGFNNATANQY 150
|||||

RESULT 3
Q7X240
ID      Q7X240      PRELIMINARY;      PRT;      149 AA.
AC      Q7X240;
DT      01-OCT-2003 (TReMBLrel. 25, Created)
DT      01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DT      01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE      Curlin-csgA protein.
GN      CSGA.
OS      Citrobacter freundii.
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC      Enterobacteriaceae; Citrobacter.
OX      NCBI_TaxID=546;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=Fec4;
RA      Zogaj X., Bokranz W., Nimitz M., Romling U.;
RT      "Production of Cellulose and Curli Fimbriae by Members of the Family
RT      Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL      Infect. Immun. 72:4151-4158 (2003).
DR      EMBL; AJ515701; CAD56675.1; -.
SQ      SEQUENCE 149 AA; 15260 MW; 946DD52017F648FD CRC64;

Query Match      71.3%; Score 552; DB 2; Length 149;
Best Local Similarity 73.5%; Pred. No. 7.2e-38;
Matches 111; Conservative 18; Mismatches 20; Indels 2; Gaps 1;

QY      1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGSSGGPDSTLSIYQVGSANAALQ 60
|||||

```

```

Db      1 MKLLKVAAPAAIVVSGSALAGVVPQW- GGNHHGGSGNYGPDSSLSIYQVGSNSANAALQ 58
|||||
QY      61 SDARKYDQLVTRVVTHEMAHAGGADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGG 120
|||||
Db      59 SDARKSDVTITQHGRNGAVVQGGADDTISLKOTGFQNSATIDQWNAKNSDITVQFGG 118
|||||

QY      121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
|||||
Db      119 RGNALVNQTASDSSNVLQVQVGFNNATANQH 149
|||||

RESULT 4
Q8CW63
ID      Q8CW63      PRELIMINARY;      PRT;      152 AA.
AC      Q8CW63;
DT      01-MAR-2003 (TReMBLrel. 23, Created)
DT      01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT      01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE      Major curlin subunit precursor.
GN      CSGA OR C1306.
OS      Escherichia coli O6.
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC      Enterobacteriaceae; Escherichia.
OX      NCBI_TaxID=217992;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX      MEDLINE=22388234; PubMed=12471157;
RA      Walch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA      Rasko D., Buckles E.J., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA      Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.;
RA      Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT      of uropathogenic Escherichia coli.";
RT      "Extensive mosaic structure revealed by the complete genome sequence
RL      Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).
DR      EMBL; A8016759; AAN79779.1; -.
KW      Complete proteome.
SQ      SEQUENCE 152 AA; 15064 MW; 3BA57F34C1240E83 CRC64;

Query Match      65.1%; Score 503.5; DB 16; Length 152;
Best Local Similarity 67.1%; Pred. No. 7.1e-34;
Matches 102; Conservative 20; Mismatches 29; Indels 1; Gaps 1;

QY      1 MKLLKVAAPAAIVVSGSALAGVVPQW- GGGNGHNGGSSGGPDSTLSIYQVGSANAALQ 59
|||||
Db      1 MKLLKVAAPAAIVVSGSALAGVVPQYGGGGNGHGGGNSGPNSELNIYQVGGNSALAQ 60
|||||

QY      60 QSDARKYDQLVTRVVTHEMAHAGGADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGG 119
|||||
Db      61 QADARNSDLTITQHGCGGADVQGGSDSDSIDLTQNGFGNSATIDQWNGKSDTMTVKQFG 120
|||||

QY      120 GNNALVNQTASDSSVMVRQVGFNNATANQY 151
|||||
Db      121 GGNGAAVDQTASNSNVNVTQVGFNNATAHQY 152
|||||

RESULT 5
Q7X237
ID      Q7X237      PRELIMINARY;      PRT;      150 AA.
AC      Q7X237;
DT      01-OCT-2003 (TReMBLrel. 25, Created)
DT      01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DT      01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE      Curlin-csgA protein.
GN      CSGA.
OS      Enterobacter sakazakii.
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC      Enterobacteriaceae; Enterobacter.
OX      NCBI_TaxID=28141;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=Fec39;
RA      Zogaj X., Bokranz W., Nimitz M., Romling U.;

```

```

RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RL Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RT Infect. Immun. 72:4151-4158(2003).
DR EMBL: AJ515702; CAD56678.1; -.
SQ SEQUENCE 150 AA; 15112 MW; 5D8BB2D872DF15F3 CRC64;

Query Match 55.6%; Score 430.5; DB 2; Length 150;
Best Local Similarity 58.9%; Pred. No. 6.9e-28;
Matches 89; Conservative 27; Mismatches 34; Indels 1; Gaps 1;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSSGPDSTLSIYQVGSANAALALQ 60
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1 MKFIKVAALAIIVSSGSANAGHINQ-GGWGHGHGGYGGFNFSTINLYQNGGNSALALQ 59

QY 61 SDARKYDQLVTRVVTHEMAHAGQAGDNSTIELTQNGFRNATIDQWNAKNSDITVGYGG 120
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 60 TDARNSVLNIGTGTGGGAGDVGGQSDSDSINLTQNGFNSATLDQWNSKDSVMNVSYGG 119

QY 121 NNAALVNOTASDSSVMVQVGFGNATANQY 151
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 120 LNALVDQTASNTVNTQIGFNGHATAHQY 150

RESULT 6
Q54069 PRELIMINARY; PRT; 76 AA.
AC Q54069;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE SEF17 fimbria (Fragment).
GN AGFA.
OS Salmonella enteritidis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=592;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SE30;
RA Cox J.M., Eglezos S., Woolcock J.B.;
RT "Virulence of Salmonella enteritidis in chickens correlates with
RT colony morphology and expression of SEF17 fimbriae.";
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
FR EMBL; U53207; AAA98671.1; -.
FT NON_TER 76
FT NON_TER 1
SQ SEQUENCE 76 AA; 7704 MW; 2FD5411241A7BCB1 CRC64;

Query Match 38.9%; Score 301; DB 2; Length 76;
Best Local Similarity 80.3%; Pred. No. 1.3e-17;
Matches 61; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

QY 30 GNHNGGSSGPDSTLSIYQVGSANAALALQSDARKYDQLVTRVVTHEMAHAGQAGDNST 89
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1 GNHNGGSSGPDSTLSIYQVGSANAALALQSDARKSETTIQSGYNGADVGGQAGDNST 60

QY 90 IELTQNGFRNATIDQ 105
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 61 IELTQNGFRNATIDQ 76

RESULT 7
Q593J5 PRELIMINARY; PRT; 29 AA.
AC Q593J5;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE Curlin subunit monomer (Fragment).
GN CSGA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.

```

```

OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC TRANSPOSON=Insertion sequence IS1;
RX MEDLINE=99314153; PubMed=10386375;
RA La Regione R.M., Collighan R.J., Woodward M.J.;
RT "Non-cultivation of Escherichia coli O78:H80 isolates associated with
RT ISI inserti on in csbs and reduced persistence in poultry infection.";
RL FEMS Microbiol. Lett. 175:247-253(1999).
DR EMBL; AJ131756; CAB45380.1; -.
FT NON_TER 29
FT NON_TER 29
SQ SEQUENCE 29 AA; 2789 MW; E290DFC07ABBB243 CRC64;

Query Match 15.8%; Score 122; DB 2; Length 29;
Best Local Similarity 89.7%; Pred. No. 0.0019;
Matches 26; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGG 29
   |||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1 MKLLKVAATAAIVFSGSALAGVVPQYGGG 29

RESULT 8
Q89J14 PRELIMINARY; PRT; 130 AA.
AC Q89J14;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE B115299 protein.
GN B115299.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RX MEDLINE=22484998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA Sasamoto S., Watanabe A., Idesawa K., Iriguchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimpo S., Teurtooka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110.";
RL DNA Res. 9:189-197(2002).
DR EMBL; AP005954; BAC50564.1; -.
KW Complete proteome.
SQ SEQUENCE 130 AA; 12699 MW; ACFB2D66A48D260F CRC64;

Query Match 14.9%; Score 115; DB 16; Length 130;
Best Local Similarity 25.9%; Pred. No. 0.046;
Matches 41; Conservative 27; Mismatches 48; Indels 42; Gaps 6;

QY 4 LKVAFAAIVVSGSALAGVVPQWGGG-----GNHNGG-----GNSSGPDSTLSIYQGS 52
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1 MRITYLVAATAIALSALTVDQAQAGNSASVLQFGTTNSSFSISQGTSTNNATTL---QFGA 57

QY 53 ANAALALQSDARKYDQLVTRVVTHEMAHAGQAGDNSTIELTQNGFRNATIDQWNAKNSD 112
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 58 TNTATLTQGS-----LITVNTAVTGGGTTAT-----ASNTA 90

QY 113 ITVGOYGGNNAALVNOTASDSSVMVQVGFGNATANQ 150
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 91 LT-GQVGGSSSLIGQIGANNATAGVQGLGLNGSTILQ 127

RESULT 9
Q8EIH3 PRELIMINARY; PRT; 139 AA.
AC Q8EIH3;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)

```

```

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Minor curlin subunit CsgB, putative.
GN S00866.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.
OX NCBI_TaxID=70863;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MR-1;
RX MEDLINE=22297686; PubMed=12368813;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
RA Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,
RA Mueller J., Khouri H., Gill J., Uterback T.R., McDonald L.A.,
RA Felblyum T.V., Smith H.O., Venter J.C., Nealsen K.H., Fraser C.M.;
RA "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis."
RL Nat. Biotechnol. 20:1118-1123(2002).
DR EMBL; AB015532; AAN53942.1; -.
DR TIGR; S00866; -.
KW Complete proteome.
SQ SEQUENCE 139 AA; 14811 MW; 41EC1CFA76957920 CRC64;

Query Match 14.2%; Score 110; DB 16; Length 139;
Best Local Similarity 28.3%; Pred. No. 0.13;
Matches 30; Conservative 18; Mismatches 34; Indels 24; Gaps 3;

QY 39 SPEDSTLSYQYGSANAALALQSDARKYDQVTRVVTHEMAHAGQADNSTIETQNGFR 98
DB 41 SGRLNLDLVQQTQNGIFVFS-----GSDNSAY-VTQAGND 77
QY 99 NNATIDQNAKSDITVGYGGNNAALVNOTASDSSVMVROVGFNG 144
DB 78 NISLVQICT-NNEVQLQVQGNKASITQIGNDNLVQLNGSGN 122

RESULT 10
Q8EIH4
ID Q8EIH4 PRELIMINARY; PRT; 502 AA.
AC Q8EIH4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE Conserved hypothetical protein.
GN S00865.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.
OX NCBI_TaxID=70863;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MR-1;
RX MEDLINE=22297686; PubMed=12368813;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
RA Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,
RA Mueller J., Khouri H., Gill J., Uterback T.R., McDonald L.A.,
RA Felblyum T.V., Smith H.O., Venter J.C., Nealsen K.H., Fraser C.M.;
RA "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis."
RL Nat. Biotechnol. 20:1118-1123(2002).
DR EMBL; AB015532; AAN53941.1; -.
DR TIGR; S00865; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 502 AA; 52441 MW; D08CA23D6C46B62D CRC64;

Query Match 13.8%; Score 106.5; DB 16; Length 502;
Best Local Similarity 23.1%; Pred. No. 1.2;
Matches 40; Conservative 30; Mismatches 72; Indels 31; Gaps 5;

QY 10 AAIWVGSGSALAGVVPQWGG-----GNHNG-----CGNS-----SGPD 42
DB 163 AVFRVEGDNDGDIKQYGNNNQAGLIADLSANVGNNDVSVIEQIGNNFAGAKGTAGND 222
QY 43 STLSIYQYGSANAALALQSDARKYDQVTRVVTHEMAHAGQ-CADNSTIETQNGFRN-- 99
DB 223 NSVDIVQKGNHGFVVALAGSENDISMEQESNTAYLSMTTGGDNTVDITQDGSNTV 282
QY 100 -NATIDQNAKSDITVGYGGNNAALVNOTASDSSVMVROVGFNGNATANQY 151
DB 283 GDSLIADIQDNDITIKQKGSNGAEFQVWGSNDVDLQKRGDANFATFGAY 335

RESULT 11
Q8U6N9
ID Q8U6N9 PRELIMINARY; PRT; 145 AA.
AC Q8U6N9;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein Atu4768.
GN ATU4768 OR AGR_L_228.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura Y.K., Zhou Y., Chen L., Wood G.B., Almeida N.F. Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutayavin T., Levy R., Li M.-J., McClelland E., Palmieri A., Gordon D.,
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
C58."
RL Science 294:2317-2323(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houmel K., Gordon J., Vaudin M., Iarchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Doughty D., Scott C., Iappas C., Markelz B.,
RA Flanagan C., Crowell C., Gurson J., Iono C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58."
RL Science 294:2323-2328(2001).
DR EMBL; AB009405; AAL45562.1; -.
DR EMBL; AB008209; AAK83682.1; -.
DR PIR; AD3143; H98143.
DR PIR; H98144; H98144.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 145 AA; 14984 MW; DEDC870E1713D51A CRC64;

Query Match 13.6%; Score 105.5; DB 16; Length 145;
Best Local Similarity 23.2%; Pred. No. 0.32;
Matches 36; Conservative 27; Mismatches 61; Indels 31; Gaps 4;

QY 3 LKVAFAAIVVSGSALAGVVPQWGG-----GNHNGSGNSGPDSTLSIYQYGSANA 55
DB 1 MIRKSFIAAGVALVGLSAAAPAMANDVRIEQYGSNAGGAQEGYGNRIPTYQNGYNR 60

```

QY 56 ALALQSDARKYDQLVTRVVTTHMAHAGQADNSTIELTQNGFRNNATIDQWNAKNSDITV 115
DB 61 IVGHQ-----YGR-----HNL SAVGEHGDHNVGSTTQGNRVAGI----- 96
QY 116 GQYGGNNAALVYNQASDSSVVRVQVFGNNATANQ 150
DB 97 GQFGSNHTTILTDGNGNIAAGVQVGRGCSANVSQ 131

RESULT 12

Q89J16 ID Q89J16 PRELIMINARY; PRT; 153 AA.
AC Q89J16;
DT 01-JUN-2003 (TRENBLrel. 24, Created)
DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE CSGB protein
GN CSGB OR BL15297.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RX MEDLINE=22484998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA Sasamoto S., Watanabe A., Idesawa K., Iriguchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimpo S., Tsuruoka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110.";
RL DNA Res. 9:189-197(2002).
DR EMBL; AP005954; BAC05052.1; -
KW Complete proteome.
SQ SEQUENCE 153 AA; 15991 MW; 4CE71DEAC375145B CRC64;

Query Match 13.6%; Score 105.5; DB 16; Length 153;
Best Local Similarity 25.7%; Pred. No. 0.34;
Matches 39; Conservative 35; Mismatches 55; Indels 23; Gaps 7;
QY 2 KLLKVAA-FAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTL-SIVQVGSANAALAL 59
DB 10 RVLAVALLAALGAATQASASIQI-----SVTPNVSEITIVQFGNDVQPVTI 58
QY 60 QSDARKYDQLVTRVVTTHMAHAGQADNSTIELTQNGFRNNATIDQWNAKNSDITVQYQ 119
DB 59 BENSERV---IARVI---QIGSGTVDATI--IQNGTRNYANVIQMGG-TTNAAVGQSG 108
QY 120 GNNALVNOTASDSSVVRVQVFGNNATANQY 151
DB 109 LSNATIDITQIGNSTNALLQIGDMNSGAVRQF 140

RESULT 13

Q9SAF2 ID Q9SAF2 PRELIMINARY; PRT; 573 AA.
AC Q9SAF2;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE F3F19.21 protein (Hypothetical protein).
GN F3F19.21 OR AT1G13190, F3F19.21 OR AT1G13190.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Vysotskaia V.S., Schwartz J.R., Yu G., Toriumi M., Lee J.M., Lenz C.,
RA Liu S., Li J., Kremenetskaia I., Lueros J., Ngan I., Gonzalez A.,

RA Altafi H., Araujo R., Chao Q., Conn L., Conway A.B., Dunn P.,
RA Kansen R., Huijar L., Kim C., Palm C., Rowley D., Shinn P., Walker M.,
RA Davis R.W., Ecker J.R., Federspiel N.A., Theologis A.;
RT "Arabidopsis thaliana chromosome 1 BAC F3F19 sequence.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Theologis A.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Onodera C.S., Quach H.L., Sakurai T.,
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Nguyen M., Southwick A., Tripp M., Palm C.J., Jones T., Wu T.,
RA Carninci P., Chen H., Cheuk R., Chan M.M., Chang C.H., Dale J.M.,
RA Deng J.M., Hayashizaki Y., Hsuan V.W., Lee J.M., Ishida J., Kamiya A.,
RA Kawai J., Kim C.J., Narusaka M., Onodera C.S., Quach H.L., Sakurai T.,
RA Satou M., Seki M., Shinn P., Tang C.C., Toriumi M., Wong C., Wu H.C.,
RA Yamada K., Yu G., Yuan S., Shinozaki K., Ecker J., Theologis A.,
RA Davis R.W.;
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA EMBL; AC007357; AAD31072.1; -
DR EMBL; AY062527; AAL32605.1; -
DR EMBL; BT002575; AAO00935.1; -
DR PIR; C86266; C86286.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 1.
DR SMART; SM00360; RRM; 1.
DR PROSITE; PS50102; RRM; 1.
KW Hypothetical protein.
SQ SEQUENCE 573 AA; 61400 MW; 0D1C10384B86C2BC CRC64;
Query Match 12.8%; Score 99; DB 10; Length 573;
Best Local Similarity 25.4%; Pred. No. 5.9;
Matches 34; Conservative 24; Mismatches 46; Indels 30; Gaps 6;
QY 13 VVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIVQVGSANAALQSDARKYDQLVTR 72
DB 84 IVSGGTVEG---KYRNDGGHNG---ISGPDTRSDVYPOASSFGAKGLNID----- 127
QY 73 VVTHEMAHAGQADNSTIELTQNGFRNNA-TIDQWNAKNS-----DITVQYVGNN 122
DB 128 IQSNKTAQOG---STTVVLNNGHFGSGNAVVPMPVTHNSYGAPQGAQOIPIVSQMSVNP 183
QY 123 AALVNOTASDSSVM 136
DB 184 NVAMNKSPQTQSFFV 197
RESULT 14
Q89J15 ID Q89J15 PRELIMINARY; PRT; 154 AA.
AC Q89J15;
DT 01-JUN-2003 (TRENBLrel. 24, Created)
DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE BL15298 protein.
GN BL15298.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.

Search completed: August 2, 2004, 14:54:39
Job time : 30.7 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:35:42 ; Search time 44.9 Seconds
(without alignments)
950.215 Million cell updates/sec

Title: US-09-543-407-22

Perfect score: 776

Sequence: 1 MKLLKVAFAAIVVSGSLA.....DSSVMVRQVGFNNATANQV 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04.*

1: Geneseq1980s.*

2: Geneseq1990s.*

3: Geneseq2000s.*

4: Geneseq2001s.*

5: Geneseq2002s.*

6: Geneseq2003as.*

7: Geneseq2003bs.*

8: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	776	100.0	151	3	AAB36351 Agfa::PT3
2	712	91.8	151	3	AAB36354 Agfa::PT3
3	682	87.9	151	2	AAR74625 Agfa sequ
4	682	87.9	151	3	AAB36341 Salmonell
5	677	87.2	151	2	Aaw23570 Salmonell
6	657	84.7	151	3	AAB36355 Agfa::PT3
7	619	79.8	151	3	AAB36350 Agfa::PT3
8	609	78.5	151	3	AAB36352 Agfa::PT3
9	604	77.8	151	3	AAB36346 Agfa::PT3
10	602	77.6	151	3	AAB36347 Agfa::PT3
11	601	77.4	151	3	AAB36353 Agfa::PT3
12	600	77.3	151	3	AAB36349 Agfa::PT3
13	567	73.1	151	3	AAB36348 Agfa::PT3
14	521	67.1	151	3	AAB36343 Escherich
15	516	66.5	151	7	ABR82651 E. coli C
16	497	64.0	120	2	AAR62761 Agfa sequ
17	497	64.0	120	2	Aaw23569 Salmonell
18	443	57.1	142	2	AAR52664 Fibronect
19	371	47.8	122	2	AAR52663 FNB curli
20	146	18.8	45	3	AAB36316 Salmonell
21	132	17.0	22	3	AAB36318 Salmonell
22	115	14.8	22	3	AAB36325 Salmonell
23	115	14.8	22	3	AAB36339 Salmonell
24	115	14.8	22	3	AAB36320 Salmonell
25	113	14.6	24	7	ABR82644 E. coli c

26	111	14.3	22	3	AAB36322	Aab36322 Salmonell
27	111	14.3	22	3	AAB36327	Aab36327 Salmonell
28	111	14.3	22	3	AAB36337	Aab36337 Salmonell
29	109	14.0	23	3	AAB36340	Aab36340 Salmonell
30	109	14.0	23	3	AAB36324	Aab36324 Salmonell
31	109	14.0	23	3	AAB36319	Aab36319 Salmonell
32	102	13.1	26	7	ABR82649	E. coli V
33	98	12.6	26	7	ABR82645	E. coli C
34	97.5	12.6	520	6	AAO16497	Aao16497 Argiote t
35	96	12.4	19	3	AAB36323	Salmonell
36	96	12.4	19	3	AAB36336	Salmonell
37	96	12.4	19	3	AAB36328	Salmonell
38	96	12.4	2309	4	ABB66232	Drosophi
39	95.5	12.3	252	8	AD83865	Chemokine
40	95	12.2	597	4	AAU08231	Polypepti
41	94.5	12.2	738	2	AAW56163	New DNA s
42	92.5	11.9	151	3	AAB36344	Escherich
43	92.5	11.9	251	5	AB245119	Human BLY
44	91	11.7	1397	7	ADD42761	Chlamydia
45	91	11.7	1751	5	ABG91039	Chlamydia

ALIGNMENTS

RESULT 1

AAB36351

ID AAB36351 standard; protein; 151 AA.

XX

AC AAB36351;

XX

DT 26-FEB-2001 (first entry)

XX

DE Agfa::PT3#6 amino acid sequence SEQ ID NO:22.

XX

XX Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;

XX

OS Salmonella enteritidis.

OS

OS Escherichia coli.

OS

XX Synthetic.

XX

PN WO2000060102-A2.

XX

PD 12-OCT-2000.

XX

PF 05-APR-2000; 2000WO-CA0000356.

XX

PR 05-APR-1999; 99US-0127888P.

XX

PA (UYVI-) UNIV VICTORIA.

XX

PI White AP, Doran JL, Collison SK, Kay WW;

XX

DR WPI; 2000-672631/65.

XX

DR N-PSDB; AAC64627.

XX

PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence

PT which encodes foreign epitope or antigen, expresses recombinant Agfa

PT protein useful for eliciting immune response in animal.

XX

PS Disclosure; Page 137; 139pp; English.

XX

CC The present invention describes a recombinant agfa gene (1) where a

CC segment of the gene has been replaced by a segment of a foreign DNA

CC sequence which encodes a foreign epitope or antigen. Also described are:

CC (1) use of thin aggregative fimbriae (SEF17/TAIF) nucleation depended

CC assembly system of strains of Salmonella, Escherichia coli and

CC Enterobacteriaceae for the production of fimbriae comprising recombinant

CC Agfa, CsgA and Agfa-homologue fimbrian subunits, respectively; (2)

CC directing recombination of a recombinant gene into the chromosome of the

CC homologous species; (3) directing recombination of a recombinant gene

CC back into the chromosome of the homologous species, replacing the native

CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fibrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;

Query Match 100.0%; Score 776; DB 3; Length 151;
 Best Local Similarity 100.0%; Pred. No. 9.5e-68;
 Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGSPDSTLSIYQGSANAALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGSPDSTLSIYQGSANAALQ 60
 QY 61 SPARKSETTITQSGYNGADVGQGDNDYDQLVTRVVTTHMAHADQNAKNSDIITVGQYGG 120
 DB 61 SPARKSETTITQSGYNGADVGQGDNDYDQLVTRVVTTHMAHADQNAKNSDIITVGQYGG 120
 QY 121 NNAALVNQTASDSVVMVRQVGFNNATANQY 151
 DB 121 NNAALVNQTASDSVVMVRQVGFNNATANQY 151

RESULT 2
 AAB36354
 ID AAB36354 standard; protein; 151 AA.
 AC AAB36354;
 XX
 XX 26-FEB-2001 (first entry)
 XX
 DE Agfa::PT3#9 amino acid sequence SEQ ID NO:28.
 XX
 KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 PN WO2000060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 FI White AP, Doran JL, Collison SK, Kay WW;
 XX
 DR WPI: 2000-672631/65.
 DR N-PSDB; AAC64630.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 138; 139pp; English.

XX The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal, Agfa
 CC comprising separating an amino acid polymer comprising a recombinant amino
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fibrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;

Query Match 91.8%; Score 712; DB 3; Length 151;
 Best Local Similarity 91.1%; Pred. No. 1.7e-61;
 Matches 144; Conservative 0; Mismatches 0; Indels 14; Gaps 2;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGSPDSTLSIYQGSANAALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGSPDSTLSIYQGSANAALQ 60
 QY 61 SPARKSETTITQSGYNGADVGQGDNDYDQLVTRVVTTHMAHA-----DOWNAKNSDI 113
 DB 61 SPARKSETTITQSGYNGAD-----YDQLVTRVVTTHMAHAFRNATIDOWNAKNSDI 113
 QY 114 TVGQYGGNNAALVNQTASDSVVMVRQVGFNNATANQY 151
 DB 114 TVGQYGGNNAALVNQTASDSVVMVRQVGFNNATANQY 151

RESULT 3
 AAR74625
 ID AAR74625 standard; protein; 151 AA.
 XX
 AC AAR74625;
 XX
 XX 25-MAR-2003 (revised)
 DT 26-JUN-1995 (first entry)
 DE Agfa sequence.
 XX
 KW Salmonella; Agfa; vaccine.
 XX
 OS Salmonella.
 XX
 PN WO9425598-A2.
 XX
 PD 10-NOV-1994.
 XX
 PF 26-APR-1994; 94WO-IB000207.
 XX
 PR 26-APR-1993; 93US-00054452.
 XX
 PA (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
 PA (KING/) KING J.
 XX

PI Kay WW, Collinson SK, Clouthier SC, Doran JL;
 XX WPI; 1994-358275/44.
 DR N-PSDB; AAQ87467.
 XX
 PT Eliciting an immune response to Salmonella - using attenuated Salmonella
 PT strains, vector constructs, or compens. contg. fimbrial type proteins.
 XX
 XX Disclosure; Fig 7B; 95pp; English.
 XX
 CC The Salmonella Agfa protein and DNA are used in vaccine and genetic
 CC immunization compositions, respectively, to elicit an immune response to
 CC Salmonella in animals (e.g. food producing animals) and humans. (Updated
 CC on 25-MAR-2003 to correct PN field.)
 XX
 XX Sequence 151 AA;
 SQ

Query Match 87.9%; Score 682; DB 2; Length 151;
 Best Local Similarity 90.7%; Pred. No. 1.4e-58;
 Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

QY 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDPSTLSIYQYGSANAALALQ 60
 DB 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDPSTLSIYQYGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGADYVGQADNVDQLVTRVVTHEMAHADQWNAKNSDITVGYGG 120
 DB 61 SDARKSETTITQSGYNGADYVGQADNVDQLVTRVVTHEMAHADQWNAKNSDITVGYGG 120

QY 121 NNAALVNTASDSSVMVRQVGFNNATANQY 151
 DB 121 NNAALVNTASDSSVMVRQVGFNNATANQY 151

RESULT 4
 AAB36341
 ID AAB36341 standard; protein; 151 AA.
 AC AAB36341;
 XX
 XX 26-FEB-2001 (first entry)
 DT
 DE Salmonella enteritidis Agfa amino acid sequence SEQ ID NO:5.
 XX
 XX Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
 KW vaccine; immune response; immunogen.
 XX
 XX Salmonella enteritidis.
 OS
 XX WO200060102-A2.
 PN
 XX 12-OCT-2000.
 PD
 XX 05-APR-2000; 2000WO-CA000356.
 PF
 XX 05-APR-1999; 99US-0127888P.
 PR
 XX (UUVI-) UNIV VICTORIA.
 XX
 XX White AP, Doran JL, Collinson SK, Kay WW;
 PI
 XX WPI: 2000-672631/65.
 DR
 XX N-PSDB; AAC64617.
 DR
 XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 XX Disclosure; Page 135; 139pp; English.
 PS
 XX The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:

CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 XX Sequence 151 AA;
 SQ

Query Match 87.9%; Score 682; DB 3; Length 151;
 Best Local Similarity 90.7%; Pred. No. 1.4e-58;
 Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

QY 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDPSTLSIYQYGSANAALALQ 60
 DB 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDPSTLSIYQYGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGADYVGQADNVDQLVTRVVTHEMAHADQWNAKNSDITVGYGG 120
 DB 61 SDARKSETTITQSGYNGADYVGQADNVDQLVTRVVTHEMAHADQWNAKNSDITVGYGG 120

QY 121 NNAALVNTASDSSVMVRQVGFNNATANQY 151
 DB 121 NNAALVNTASDSSVMVRQVGFNNATANQY 151

RESULT 5
 AAW23570
 ID AAW23570 standard; protein; 151 AA.
 XX
 XX AAW23570;
 XX
 XX 25-MAR-2003 (revised)
 DT
 DT 29-SEP-1997 (first entry)
 DT
 XX Salmonella enteritidis 27655-3b agfa.
 XX
 XX Enteropathogenic bacteria; enterobacteria; S. enteritidis; antibody.
 KW
 XX Salmonella enteritidis.
 OS
 XX Key Location/Qualifiers
 FH
 FT Misc-difference 123
 FT /note= "Encoded by GCC"
 FT
 XX US5635617-A.
 PN
 XX 03-JUN-1997.
 PD
 XX 26-APR-1994; 94US-00233788.
 PF
 XX 26-APR-1993; 93US-00054452.
 PR
 XX (UUVI-) UNIV VICTORIA INNOVATION & DEV CORP.
 PA
 XX Collinson SK, Kay WW, Doran JL;
 PI


```

PR 05-APR-1999; 99US-0127888P.
PA (UYVI-) UNIV VICTORIA.
XX
XX White AP, Doran JL, Collison SK, Kay WW;
XX WPI; 2000-672631/65.
XX N-PSDB; AAC64626.
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
XX which encodes foreign epitope or antigen, expresses recombinant Agfa
XX protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 137; 139pp; English.
XX
XX The present invention describes a recombinant agfa gene (I) where a
XX segment of the gene has been replaced by a segment of a foreign DNA
XX sequence which encodes a foreign epitope or antigen. Also described are:
XX (1) use of thin aggregative fimbriae (SEF1/TAFF) nucleation depended
XX assembly system of strains of Salmonella, Escherichia coli and
XX Enterobacteriaceae for the production of fimbriae comprising recombinant
XX Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
XX directing recombination of a recombinant gene into the chromosome of the
XX homologous species; (3) directing recombination of a recombinant gene
XX back into the chromosome of the homologous species, replacing the native
XX copy of that gene; and (4) eliciting an immune response in an animal,
XX comprising separating an amino acid polymer comprising a recombinant Agfa
XX protein containing a replacement segment or segments of foreign amino
XX acid sequence or sequences grown on a Salmonella, E. coli or
XX Enterobacteriaceae host cell, from the host cell and introducing the
XX polymer into the animal in conjunction with a carrier or diluent. (1) is
XX useful for the expression of recombinant Agfa protein which is useful for
XX eliciting an immune response in an animal. In a fimbrial presentation
XX system the heterologous antigens are presented in high numbers (up to
XX 500,000 copies/cell), the hybrid fimbrial protein possesses both the
XX immunogenicity and adhesion properties relevant for an efficient live
XX vaccine, the carrier fimbrial subunit proteins are usually strong
XX immunogens, which may be important for directing an immune response
XX against the inserted epitope, and hybrid fimbriae are easy and
XX inexpensive to purify in large amount. The present sequence is given in
XX the exemplification of the present invention
XX
XX Sequence 151 AA;
XX
XX Query Match 79.8%; Score 619; DB 3; Length 151;
XX Best Local Similarity 74.6%; Pred. No. 2e-52; Indels 44; Gaps 2;
XX Matches 129; Conservative 0; Mismatches 0;
XX
XX 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGDPDSTLSIYQYGSANAALQ 60
XX
XX 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGDPDSTLSIYQYGSANAALQ 60
XX
XX 61 SDARKSETTITQSGYNGADVGQGDNDYDQLVTRVVTHEMAHA----- 103
XX
XX 61 SDARK-----YDQLVTRVVTHEMAHAGQGDNDSTIELTONGFR 98
XX
XX 104 -----DOWNAKNSDITVQYGGNNAALVNQTASDSSVVMVQVGFGNNTANQY 151
XX
XX 99 NNATIDQWNAKNSDITVQYGGNNAALVNQTASDSSVVMVQVGFGNNTANQY 151
XX
XX
XX RESULT 8
XX AAB36352
XX ID AAB36352 standard; protein; 151 AA.
XX
XX XX AAB36352;
XX
XX XX
XX XX 26-FEB-2001 (first entry)
XX DT
XX DE Agfa::PT3#7 amino acid sequence SEQ ID NO:24.
XX DE
XX KW Salmonella; agfa; chromosomal gene replacement; fimbrial; epitope;
XX vaccine; immune response; immunogen.

```

```

XX
XX Salmonella enteritidis.
XX Escherichia coli.
XX Synthetic.
XX
XX WO200060102-A2.
XX
XX 12-OCT-2000.
XX
XX 05-APR-2000; 2000WO-CA000356.
XX
XX 05-APR-1999; 99US-0127888P.
XX
XX (UYVI-) UNIV VICTORIA.
XX
XX White AP, Doran JL, Collison SK, Kay WW;
XX WPI; 2000-672631/65.
XX N-PSDB; AAC64628.
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
XX which encodes foreign epitope or antigen, expresses recombinant Agfa
XX protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 138; 139pp; English.
XX
XX The present invention describes a recombinant agfa gene (I) where a
XX segment of the gene has been replaced by a segment of a foreign DNA
XX sequence which encodes a foreign epitope or antigen. Also described are:
XX (1) use of thin aggregative fimbriae (SEF1/TAFF) nucleation depended
XX assembly system of strains of Salmonella, Escherichia coli and
XX Enterobacteriaceae for the production of fimbriae comprising recombinant
XX Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
XX directing recombination of a recombinant gene into the chromosome of the
XX homologous species; (3) directing recombination of a recombinant gene
XX back into the chromosome of the homologous species, replacing the native
XX copy of that gene; and (4) eliciting an immune response in an animal,
XX comprising separating an amino acid polymer comprising a recombinant Agfa
XX protein containing a replacement segment or segments of foreign amino
XX acid sequence or sequences grown on a Salmonella, E. coli or
XX Enterobacteriaceae host cell, from the host cell and introducing the
XX polymer into the animal in conjunction with a carrier or diluent. (1) is
XX useful for the expression of recombinant Agfa protein which is useful for
XX eliciting an immune response in an animal. In a fimbrial presentation
XX system the heterologous antigens are presented in high numbers (up to
XX 500,000 copies/cell), the hybrid fimbrial protein possesses both the
XX immunogenicity and adhesion properties relevant for an efficient live
XX vaccine, the carrier fimbrial subunit proteins are usually strong
XX immunogens, which may be important for directing an immune response
XX against the inserted epitope, and hybrid fimbriae are easy and
XX inexpensive to purify in large amount. The present sequence is given in
XX the exemplification of the present invention
XX
XX Sequence 151 AA;
XX
XX Query Match 78.5%; Score 609; DB 3; Length 151;
XX Best Local Similarity 73.6%; Pred. No. 1.9e-51;
XX Matches 128; Conservative 0; Mismatches 0; Indels 46; Gaps 2;
XX
XX 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGDPDSTLSIYQYGSANAALQ 60
XX
XX 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGDPDSTLSIYQYGSANAALQ 60
XX
XX 61 SDARKSETTITQSGYNGADVGQGDNDYDQLVTRVVTHEMAHA----- 97
XX
XX 61 SDARKSETTITQSGYNGADVGQGDNDSTIELTONGFRNNATIDQWNAKNSDITVQV 120
XX
XX 98 HEMAHADQWNAKNSDITVQYGGNNAALVNQTASDSSVVMVQVGFGNNTANQY 151
XX
XX 121 HEMAH-----NOTASDSSVVMVQVGFGNNTANQY 151
XX
XX
XX RESULT 9

```

AAB36346
 ID AAB36346 standard; protein; 151 AA.
 AC AAB36346;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Agfa::PT3#1 amino acid sequence SEQ ID NO:12.
 XX
 KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
 XX vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 PN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay WW;
 XX
 DR WPI; 2000-672631/65.
 DR N-PSDB; AAC64622.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 135; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrin subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;
 Query Match 77.8%; Score 604; DB 3; Length 151;
 Best Local Similarity 80.8%; Pred. No. 5.9e-51;
 Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;
 1 MKLLKVAFAAATVWSGSLAGYVVPQWGGGNGHGGSSGPDSTLSIYQYGSANAALQ 60
 1 MKLLKVAFAAATVWSGSLAGYVVPQWGGGNGHGGSSGPDSTLSIYQYGSANAALQ 60

61 SDARKSETTITQSGYNGADVGGADNYDOLVTRVVTHEMAHADOWNAKNSDITVQYGG 120
 ||||| :|||
 61 SDARKSETTITQSGYNGADVGGADNSTIELTQNGFRNATIDOWNAKNSDITVQYGG 120
 ||||| :|||
 121 NNAALVNYQASDSSVMVRQVGFONNATANQY 151
 ||||| :|||
 121 NNAALVNYDOLVTRVVTHEMAHANNATANQY 151
 ||||| :|||

RESULT 10
 AAB36347
 ID AAB36347 standard; protein; 151 AA.
 AC AAB36347;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Agfa::PT3#2 amino acid sequence SEQ ID NO:14.
 XX
 KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
 XX vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 PN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay WW;
 XX
 DR WPI; 2000-672631/65.
 DR N-PSDB; AAC64623.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 136; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrin subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;
 Query Match 77.8%; Score 604; DB 3; Length 151;
 Best Local Similarity 80.8%; Pred. No. 5.9e-51;
 Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;
 1 MKLLKVAFAAATVWSGSLAGYVVPQWGGGNGHGGSSGPDSTLSIYQYGSANAALQ 60
 1 MKLLKVAFAAATVWSGSLAGYVVPQWGGGNGHGGSSGPDSTLSIYQYGSANAALQ 60

CC the exemplification of the present invention
 XX Sequence 151 AA;
 SQ

Query Match 77.6%; Score 602; DB 3; Length 151;
 Best Local Similarity 81.5%; Pred. No. 9.3e-51;
 Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;

Qy 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
 Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60

Qy 61 SDARKSETTITQSGYNGADYDQGVGGADNYDQVTRVVTHEMAHADQWNAKNSDITVQYGG 120
 Db 61 SDARKSETTITQSGYNGADYDQGVGGADNYDQVTRVVTHEMAHADQWNAKNSDITVQYDQ 120

Qy 121 NNAALVNQTSADSSVVRQVFGNNATANQY 151
 Db 121 LVTRVVTHEMAHAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60

RESULT 11
 AAB36353
 ID AAB36353 standard; protein; 151 AA.
 AC AAB36353;
 DT 26-FEB-2001 (first entry)
 XX AgfA::PT3#8 amino acid sequence SEQ ID NO:26.
 KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
 KW vaccine; immune response; immunogen.
 XX Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX WO2000060102-A2.
 XX 12-OCT-2000.
 XX 05-APR-2000; 2000WO-CA000356.
 XX 05-APR-1999; 99US-0127888P.
 XX (UYVI-) UNIV VICTORIA.
 XX White AP, Doran JL, Collison SK, Kay WW;
 XX WPI; 2000-672631/65.
 XX N-PSDB; AAC64629.
 XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant AgfA
 PT protein useful for eliciting immune response in animal.
 XX Disclosure; Page 138; 139pp; English.

CC The present invention describes a recombinant agfA gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAf) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant AgfA
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or

CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX Sequence 151 AA;
 SQ

Query Match 77.4%; Score 601; DB 3; Length 151;
 Best Local Similarity 81.5%; Pred. No. 1.2e-50;
 Matches 123; Conservative 4; Mismatches 24; Indels 0; Gaps 0;

Qy 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
 Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60

Qy 61 SDARKSETTITQSGYNGADYDQGVGGADNYDQVTRVVTHEMAHADQWNAKNSDITVQYGG 120
 Db 61 LVTRVVTHEMAHAGYNGADYDQGVGGADNYDQVTRVVTHEMAHADQWNAKNSDITVQYGG 120

Qy 121 NNAALVNQTSADSSVVRQVFGNNATANQY 151
 Db 121 NNAALVNQTSADSSVVRQVFGNNATANQY 151

RESULT 12
 AAB36349
 ID AAB36349 standard; protein; 151 AA.
 XX AAB36349;
 AC AAB36349;
 XX 26-FEB-2001 (first entry)
 XX AgfA::PT3#4 amino acid sequence SEQ ID NO:18.
 XX Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
 KW vaccine; immune response; immunogen.
 XX Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX WO2000060102-A2.
 XX 12-OCT-2000.
 XX 05-APR-2000; 2000WO-CA000356.
 XX 05-APR-1999; 99US-0127888P.
 XX (UYVI-) UNIV VICTORIA.
 XX White AP, Doran JL, Collison SK, Kay WW;
 XX WPI; 2000-672631/65.
 XX N-PSDB; AAC64625.
 XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant AgfA
 PT protein useful for eliciting immune response in animal.
 XX Disclosure; Page 136; 139pp; English.

CC The present invention describes a recombinant agfA gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAf) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant AgfA
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or

CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant AgfA
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX Sequence 151 AA;

Query Match 77.3%; Score 600; DB 3; Length 151;
 Best Local Similarity 81.5%; Pred. No. 1.5e-50;
 Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNSGPDSTLSIYQGSANAALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNSGPDSTLSIYQGSANAALQ 60
 QY 61 SPARKSETTITQSGYNGADVGGADNYDQLVTRVVTTHMAHADOWNAKNSDITVGOYGG 120
 DB 61 SPARKSETTITQSGYNGADVGGADNSTIETQTQGFNRNATIDOWNAKNSDITVGOYGG 120
 QY 121 NNAALVNQTADSSVMVRQVGFNNATANQY 151
 DB 121 NNAALVNQTADSSVMVRQVGFNNATANQY 151

RESULT 13
 ID AAB36348
 XX AAB36348 standard; protein; 151 AA.
 AC AAB36348;
 XX AAB36348;
 DT 26-FEB-2001 (first entry)
 XX AgfA::PT3#3 amino acid sequence SEQ ID NO:16.
 DE AgfA::PT3#3 amino acid sequence SEQ ID NO:16.
 XX Salmonella; agfA; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 XX Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX WO200060102-A2.
 PN 12-OCT-2000.
 PD 12-OCT-2000.
 XX 05-APR-2000; 2000WO-CA000356.
 PF 05-APR-1999; 99US-0127889P.
 PR (UYVI-) UNIV VICTORIA.
 XX White AP, Doran JL, Collison SK, Kay WW;
 PI WPI; 2000-672631/65.

DR N-PSDB; AAC64624.
 XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant AgfA
 PT protein useful for eliciting immune response in animal.
 XX Disclosure; Page 136; 139pp; English.
 CC The present invention describes a recombinant agfA gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended are:
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant AgfA
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX Sequence 151 AA;

Query Match 73.1%; Score 567; DB 3; Length 151;
 Best Local Similarity 80.8%; Pred. No. 2.4e-47;
 Matches 122; Conservative 5; Mismatches 24; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNSGPDSTLSIYQGSANAALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNSGPDSTLSIYQGSANAALQ 60
 QY 61 SPARKSETTITQSGYNGADVGGADNYDQLVTRVVTTHMAHADOWNAKNSDITVGOYGG 120
 DB 61 SPARKSETTITQSGYNGADVGGADNSTIETQTQGFNRNATIDOWNAKNSDITVGOYGG 120
 QY 121 NNAALVNQTADSSVMVRQVGFNNATANQY 151
 DB 121 NNAALVNQTADSSVMVRQVGFNNATANQY 151

RESULT 14
 ID AAB36343
 XX AAB36343 standard; protein; 151 AA.
 AC AAB36343;
 XX AAB36343;
 DT 26-FEB-2001 (first entry)
 XX Escherichia coli CsgA amino acid sequence SEQ ID NO:7.
 DE Escherichia coli CsgA amino acid sequence SEQ ID NO:7.
 XX Salmonella; agfA; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 XX Escherichia coli.
 OS WO200060102-A2.
 PN 12-OCT-2000.
 PD 12-OCT-2000.
 XX


```
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UVTI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR N-PSDB; AAC64619.
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 135; 139pp; English.
XX
XX The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/IAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 151 AA;
Query Match 67.1%; Score 521; DB 3; Length 151;
Best Local Similarity 69.5%; Pred. No. 7.5e-43;
Matches 105; Conservative 17; Mismatches 29; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAIAAIVFSGSALAGVVPQYGGGNGHGGNGNSGPNSEINIIYQYGGNSALALQ 60
QY 61 SDARKSETTITQSGYNGADYVCGADNYDQIVTRVVTHEMAHADOWNAKNSDITVGYGG 120
DB 61 TDARNSDLITITQHGCGGADYVCGGSDSDSIDLTQKFGNSATLDQWNGKNSMTVKQFGG 120
QY 121 NNAALVNOTASDSSVMVVRQVGFGNNAATANY 151
DB 121 GNGAAVDQTASNSSVNVITQVGFGNNAATANY 151
RESULT 15
ABR82651
XX ABR82651 standard; protein; 151 AA.
XX
AC ABR82651;
XX
DT 04-DEC-2003 (first entry)
XX
DE E. coli CsgA subunit 15 kDa protein.
XX
```

```
KW Plasma protein; immune response; antibacterial; vaccine; gene therapy.
XX Escherichia coli.
XX
XX WO2003064446-A2.
XX
XX 07-AUG-2003.
XX
XX 30-JAN-2003; 2003WO-EP000943.
XX
XX 31-JAN-2002; 2002GB-00002275.
XX (HANS-) HANSA MEDICAL RES AB.
XX
XX Bjoerck L, Olsen A, Wikstroem M, Herwald H;
XX
XX WPI; 2003-646136/61.
XX N-PSDB; ACP36153.
XX
XX New isolated peptide capable of binding a mammalian plasma protein,
XX useful in the manufacture of a medicament for the prevention and/or
XX treatment of a bacterial infection, such as Escherichia coli, Salmonella
XX or Shigella infections.
XX
XX Disclosure; Page 41-42; 42pp; English.
XX
XX The invention relates to an isolated peptide capable of binding a
XX mammalian plasma protein or of generating an immune response in a mammal
XX selected from sequences shown in ABR82642, ABR82648-49. The peptide or
XX antibody is useful for treating a bacterial infection in a human or
XX animal or in the manufacture of a medicament for the prophylactic
XX treatment of a bacterial infection, such as Escherichia coli, Salmonella
XX or Shigella infection. The peptide that is immobilized on a solid support
XX is also useful as a reagent for determining the ability of a plasma
XX protein to bind to bacteria. The present sequence represents an E. coli
XX 15 kDa protein
XX
SQ Sequence 151 AA;
Query Match 66.5%; Score 516; DB 7; Length 151;
Best Local Similarity 68.9%; Pred. No. 2.3e-42;
Matches 104; Conservative 17; Mismatches 30; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVEAIAAIVFSGSALAGVVPQYGGGNGHGGGNGNSGPNSEINIIYQYGGNSALALQ 60
QY 61 SDARKSETTITQSGYNGADYVCGADNYDQIVTRVVTHEMAHADOWNAKNSDITVGYGG 120
DB 61 TDARNSDLITITQHGCGGADYVCGGSDSDSIDLTQKFGNSATLDQWNGKNSMTVKQFGG 120
QY 121 NNAALVNOTASDSSVMVVRQVGFGNNAATANY 151
DB 121 GNGAAVDQTASNSSVNVITQVGFGNNAATANY 151
Search completed: August 2, 2004, 14:48:27
Job time : 45.9 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (C) 1993 - 2004 Compugen Ltd.

QM protein - protein search, using sw model

Run on: August 2, 2004, 14:40:48 ; Search time 12 seconds
(without alignments)
649.627 Million cell updates/sec

Title: US-09-543-407-22
Perfect score: 776
Sequence: 1 MKLLKVAFAAIWVGSAALA.....DSSVMVRQVFGGNATANQY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/ptodata/2/iaa/5A-COMB.pcp.*
- 2: /cgn2_6/ptodata/2/iaa/5B-COMB.pcp.*
- 3: /cgn2_6/ptodata/2/iaa/6A-COMB.pcp.*
- 4: /cgn2_6/ptodata/2/iaa/6B-COMB.pcp.*
- 5: /cgn2_6/ptodata/2/iaa/PGTUS-COMB.pcp.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	677	87.2	151	1	US-08-233-788A-59
2	497	64.0	120	1	US-08-233-788A-57
3	94.5	12.2	738	3	US-08-864-038A-3
4	87	11.2	892	4	US-09-336-447A-5
5	86.5	11.1	238	4	US-09-495-880A-42
6	86.5	11.1	585	4	US-09-620-412C-337
7	86.5	11.1	585	4	US-09-598-419-337
8	86.5	11.1	1752	4	US-09-556-877-180
9	86.5	11.1	1752	4	US-09-620-412C-180
10	86.5	11.1	1752	4	US-09-598-419-180
11	85.5	11.0	589	4	US-09-498-039A-7849
12	82	10.6	975	4	US-09-328-352-4764
13	80	10.3	461	4	US-09-252-991A-24717
14	80	10.3	812	4	US-09-252-991A-30710
15	80	10.3	1034	4	US-09-252-991A-26658
16	79	10.2	943	4	US-09-056-556-204
17	79	10.2	943	4	US-09-072-596-199
18	79	10.2	943	4	US-09-477-135A-131
19	79	10.2	943	4	US-09-072-967-204
20	79	10.2	2315	4	US-09-543-681A-5434
21	78.5	10.1	209	4	US-09-125-619-42
22	78.5	10.1	278	3	US-09-260-283-2
23	78.5	10.1	309	4	US-09-252-991A-22266
24	78.5	10.1	745	4	US-09-336-115C-6
25	78	10.1	392	2	US-08-387-942C-23
26	78	10.1	553	2	US-08-387-942C-2
27	78	10.1	645	4	US-09-919-172-41
28	78	10.1	645	4	US-09-919-172-41

28	78	10.1	878	4	US-09-540-236-3401	Sequence 3401, Ap
29	78	10.1	1216	4	US-09-134-000C-5130	Sequence 5130, Ap
30	77.5	10.0	212	4	US-09-125-619-32	Sequence 32, Appl
31	77.5	10.0	437	2	US-08-737-716-2	Sequence 2, Appl
32	77.5	10.0	673	3	US-09-196-387-8	Sequence 8, Appl
33	77.5	10.0	673	4	US-08-841-835-8	Sequence 8, Appl
34	77.5	10.0	702	4	US-09-252-991A-22119	Sequence 22119, A
35	77.5	10.0	941	4	US-09-336-447A-9	Sequence 9, Appl
36	77.5	10.0	949	3	US-09-196-387-10	Sequence 10, Appl
37	77.5	10.0	949	4	US-09-841-835-10	Sequence 2, Appl
38	77.5	10.0	1327	3	US-09-196-387-2	Sequence 2, Appl
39	77.5	10.0	1327	4	US-09-841-835-2	Sequence 8, Appl
40	77.5	10.0	1327	4	US-09-972-115A-8	Sequence 2, Appl
41	77.5	10.0	1690	4	US-09-595-684B-39	Sequence 39, Appl
42	77	9.9	266	4	US-09-495-880A-26	Sequence 26, Appl
43	77	9.9	339	4	US-09-252-991A-32096	Sequence 32096, A
44	77	9.9	873	4	US-09-336-447A-13	Sequence 13, Appl
45	77	9.9	1139	1	US-08-537-210A-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-08-233-788A-59
; Sequence 59, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, Karen S.
; APPLICANT: Clouthier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; TITLE OF INVENTION: OF SALMONELLA
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,788A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERY
; INFORMATION FOR SEQ ID NO: 59:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 151 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-233-788A-59

Query Match 87.2%; Score 677; DB 1; Length 151;
Best Local Similarity 90.1%; Pred. No. 3.3e-59;
Matches 136; Conservative 2; Mismatches 13; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIWVGSAALAGVVPOWGGGNGGNSGPDSTLSIYQYGSANALAIQ 60
|||||

Db 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNHNGGSSGPDSTLSIYQYGSANAALALQ 60
 QY 61 SPARKSETTITQSGYNGADVQGGADNDVQLVTRVVTHEMAHADQWNAKNSDITVQYGG 120
 Db 61 SPARKSETTITQSGYNGADVQGGADNSTIELTQNGFRNNAITQWNAKNSDITVQYGG 120
 QY 121 NNAALVNQTSADSSVMVRQVGFNNATANQY 151
 Db 121 NNPALVNQTSADSSVMVRQVGFNNATANQY 151

RESULT 2

US-08-233-788A-57
 ; Sequence 57, Application US/08233788A
 ; Patent No. 5635617
 ; GENERAL INFORMATION:
 ; APPLICANT: Doran, James L.
 ; APPLICANT: Kay, William W.
 ; APPLICANT: Collinson, Karen S.
 ; APPLICANT: Clouthier, Sharon C.
 ; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
 ; TITLE OF INVENTION: OF SALMONELLA
 ; NUMBER OF SEQUENCES: 61
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Seed and Berry
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue
 ; CITY: Seattle
 ; STATE: Washington
 ; COUNTRY: U.S.A.
 ; ZIP: 98104-7092

COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/233,788A
 ; FILING DATE: 26-APR-1994
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: King, Joshua
 ; REGISTRATION NUMBER: 35,570
 ; REFERENCE/DOCKET NUMBER: 920043.403C2

TELEPHONE: (206) 622-4900
 TELEFAX: (206) 682-6031
 TELEX: 3723836 SEEDANBERRY

INFORMATION FOR SEQ ID NO: 57:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 120 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein

US-08-233-788A-57

Query Match 64.0%; Score 497; DB 1; Length 120;
 Best Local Similarity 87.5%; Pred. No. 1.2e-41;
 Matches 98; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

QY 22 VFPWGGGNHNGGSSGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGADV 81
 Db 1 VFPWGGGNHNGGSSGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGADV 60
 QY 82 GQADNDVQLVTRVVTHEMAHADQWNAKNSDITVQYGGNNAALVNQTSADS 133
 Db 61 GQADNSTIELTQNGFRNNAITQWNAKNSDITVQYGGNNAALVNQTSADS 112

RESULT 3

US-08-864-038A-3
 ; Sequence 3, Application US/08864038A
 ; Patent No. 6001592
 ; GENERAL INFORMATION:

APPLICANT: Kunio NAKASHIMA et al.
 ; TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE CDNA, VECTOR
 ; TITLE OF INVENTION: CONTAINING SAID CDNA, HOST CELLS TRANSFORMED WITH SAID
 ; TITLE OF INVENTION: VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING
 ; TITLE OF INVENTION: SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBODY
 ; TITLE OF INVENTION: TO SAID POLYPEPTIDE
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: 812-5 Hirano
 ; STREET: Isehinden
 ; CITY: Tsu-city
 ; STATE: Mie-prefecture
 ; COUNTRY: JAPAN
 ; ZIP: 514-01

COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: Microsoft Windows 95
 ; SOFTWARE: Word Perfect 6.1
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/864,038A
 ; FILING DATE: May 28, 1997

PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: JP 8-184459
 ; FILING DATE: 15-July-1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: C. Bruce Hamburg
 ; REGISTRATION NUMBER: 22,389
 ; REFERENCE/DOCKET NUMBER: F-5610

TELEPHONE: (212) 986-2340
 TELEFAX: (212) 953-7733
 INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 738
 ; TYPE: amino acid
 ; TOPOLOGY: linear

MOLECULE TYPE: protein
 ; ORIGINAL SOURCE:
 ; ORGANISM: Pinctada fucata
 ; CELL TYPE: mantle epithelial cell
 ; FEATURE:
 ; NAME/KEY: peptide
 ; LOCATION: from 1 to 738

IDENTIFICATION METHOD: E (by experiment)
 ; US-08-864-038A-3

Query Match 12.2%; Score 94.5; DB 3; Length 738;
 Best Local Similarity 27.6%; Pred. No. 0.48;
 Matches 43; Conservative 14; Mismatches 58; Indels 41; Gaps 6;

QY 3 LLKVAAPAAIVVSGSALAGVVPQWGGGNHNGGSSGPDSTLSIYQYGSANAALALQSD 62
 Db 419 LLKSSASASASASASAG-----GGGGGNGGNGGGG-----GGGAGALA----- 460

QY 63 ARKSETTITQSGYNGADVQGGADNDVQLVTRVVTHEMAHADQWNAKNSDITVQYCGN 121
 Db 461 -----AALAAAGAGGGLGGGGGG-----AALAAAGAGGAGGGGGLGGGLGGG 507

QY 122 NNAALVNQTSADSS-----VMVRQVGFNNATA 148
 Db 508 SAAAAAASGGGGRALRRALRRQMRGGGSAAA 543

RESULT 4

US-09-336-447A-5
 ; Sequence 5, Application US/09336447A
 ; Patent No. 6310190

GENERAL INFORMATION:
 ; APPLICANT: HANSEN, ERIC J.
 ; APPLICANT: AEBI, CHRISTOPH
 ; APPLICANT: COPE, LESLIE D.
 ; APPLICANT: MACIVER, ISOBEL

```
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USP41 AND USP42 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/09/336,447A
; CURRENT FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 5
; LENGTH: 892
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-336-447A-5

Query Match      11.2%; Score 87; DB 4; Length 892;
Best Local Similarity 26.8%; Pred. No. 3.4;
Matches 38; Conservative 15; Mismatches 57; Indels 32; Gaps 7;

QY 28 GGGNHN-----GGNNGSGPDTLSIYQYGSANAALALQSDARKSETTITQSGY--- 75
DB 89 GGGDYNEAKGNYSTVGGSSNTAKGSKSTIGGGDTN-----DANGTYSTIGGGYSRA 141
QY 76 -GNGADVQGGADNDVQLVTRVVTHEMAHADQWNAKSDITVQYG---GNNAAALV-----N 127
DB 142 IGSSTTIGGG--YTNQATGKSTVAGGRNNQATGNNSTVAGGSYNOATGNNSTVAGGSHN 199
QY 128 QTASDSSVMVQVQFGNNATAN 149
DB 200 QATGEGSF---AAGVENKANAN 218

RESULT 5
US-09-495-880A-42
; Sequence 42, Application US/09495880A
; Patent No. 6667150
; GENERAL INFORMATION:
; APPLICANT: RUDERT, FRITZ
; APPLICANT: GE, LIMING
; APPLICANT: ILAG, VIC
; TITLE OF INVENTION: NOVEL METHOD AND PHASE FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES ENCODING MEMBERS OF A MULTIMERIC
; TITLE OF INVENTION: (POLY) PEPTIDE COMPLEX
; FILE REFERENCE: MORPHO/9
; CURRENT APPLICATION NUMBER: US/09/495,880A
; CURRENT FILING DATE: 2000-02-01
; PRIOR APPLICATION NUMBER: PCT/EP98/04836
; PRIOR FILING DATE: 1998-08-03
; PRIOR APPLICATION NUMBER: EP 97 11 3319.4
; PRIOR FILING DATE: 1997-08-01
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 42
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: ompA-FLAG-peptide3-
; OTHER INFORMATION: gene IIIs encoded by phage vector fpep3_1B-IR3seq (circular)
US-09-495-880A-42

Query Match      11.1%; Score 86.5; DB 4; Length 238;
Best Local Similarity 24.3%; Pred. No. 0.69;
Matches 43; Conservative 24; Mismatches 63; Indels 47; Gaps 8;

QY 5 KVAFAAIVVSGSALAG-----VVPGWGGGNNH--NGGNGSSGPDSTLSI 47
DB 3 KTAIAIALAGFATVQAQYKVDICVYHAHLVAKCGGSGSEFNAGGSGG----- 55
QY 48 YQYGSANAALQSDARKSETTITQSGYNGADYQGGADNDVQLVTRVVTHEMAHADQWNA 107
DB 56 ---GSGGSGEGGEGGSGGEGG--GSGGSGSGGDFDYKEMANANKGAMTENADE-N 110
QY 108 AKNSDI-----TVQYQG-----GNNAAALVNQTA-----SDSSVMVQVQFGNNA 146
```

```
DB 111 ALQSDAKGLDSVATDYGAADIGFIDYVGLANGATGDFAGNSQMAQVQGDGNS 167

RESULT 6
US-09-620-412C-337
; Sequence 337, Application US/09620412C
; Patent No. 6448234
; GENERAL INFORMATION:
; APPLICANT: Steven P. Fling
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
; FILE REFERENCE: 210121.469C7
; CURRENT APPLICATION NUMBER: US/09/620,412C
; CURRENT FILING DATE: 2000-07-20
; NUMBER OF SEQ ID NOS: 363
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 337
; LENGTH: 585
; TYPE: PRT
; ORGANISM: Chlamydia trachomatis
US-09-620-412C-337

Query Match      11.1%; Score 86.5; DB 4; Length 585;
Best Local Similarity 31.6%; Pred. No. 2.2;
Matches 43; Conservative 12; Mismatches 58; Indels 23; Gaps 7;

QY 4 LKVAFAAIVVSGSALAGVVPQWGGGN--HNGGNGSSGPDSS---TLSIYQYGSANAAL 57
DB 206 LQAQASAG---NADAWASSSPQSGGATTVSDSGSSGSDTSETVPTAKGG----- 257
QY 58 ALQSDARKSETTIT---QSGYNGADYQGGADNDVQLVTRVVTHEMAHADQWNAKNSDIT 114
DB 258 GLYTDKNLSITNITGIIIEIANNAKATDVGGGA-----YVKGTLTCENSHRLQFLKNSDDQ 312
QY 115 VQY-YGGNNAALVNQOT 129
DB 313 GGGIYGEDNITLSNLT 328

RESULT 7
US-09-598-419-337
; Sequence 337, Application US/09598419
; Patent No. 6565856
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Scholler, John
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
; FILE REFERENCE: 210121.469C6
; CURRENT APPLICATION NUMBER: US/09/598,419
; CURRENT FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 357
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 337
; LENGTH: 585
; TYPE: PRT
; ORGANISM: Chlamydia trachomatis
US-09-598-419-337

Query Match      11.1%; Score 86.5; DB 4; Length 585;
Best Local Similarity 31.6%; Pred. No. 2.2;
Matches 43; Conservative 12; Mismatches 58; Indels 23; Gaps 7;

QY 4 LKVAFAAIVVSGSALAGVVPQWGGGN--HNGGNGSSGPDSS---TLSIYQYGSANAAL 57
DB 206 LQAQASAG---NADAWASSSPQSGGATTVSDSGSSGSDTSETVPTAKGG----- 257
QY 58 ALQSDARKSETTIT---QSGYNGADYQGGADNDVQLVTRVVTHEMAHADQWNAKNSDIT 114
DB 258 GLYTDKNLSITNITGIIIEIANNAKATDVGGGA-----YVKGTLTCENSHRLQFLKNSDDQ 312
QY 115 VQY-YGGNNAALVNQOT 129
```


Db 319 ACPRGQSTGSPVTTVDTAAGNLVNWKAGALYRLTEQGNVYV-----NYAISQPPGGS 372
QY 134 SVMVRQVGFGNATANQY 151
Db 373 SPALAAAGSGNSANRTDF 390

RESULT 12
US-09-328-352-4764
; Sequence 4764, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-039A
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4764
; LENGTH: 975
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-4764

Query Match 10.6%; Score 82; DB 4; Length 975;
Best Local Similarity 24.1%; Pred. No. 12;
Matches 38; Conservative 15; Mismatches 47; Indels 58; Gaps 7;
QY 15 GSGALAGVVPQGGGNGNG--GNSGDPSTLSIYQYG-----SANA 55
Db 300 AGNGIA-----SGNGEYHNGGNGGDDVDITPITGLNLSGNSFTLNGSSSSSVNT 353
QY 56 ALALOSDARKSETTI-----TQSGYG-----NGADYQGGADNYDQLVTRVWTHM 100
Db 354 APTTTSNTVNDNTIDNGSGGTGSGGNGSGDGLLNGAASNGEHNYG-----402

QY 101 AHADOWNAKNSDIT-----VGQYGGNNAALVNQTSADS 133
Db 403 --IGNGNGDDVDITSPITGIFNFGNSFSLGNSSSSS 438

RESULT 13
US-09-252-991A-24717
; Sequence 24717, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24717
; LENGTH: 461
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24717

Query Match 10.3%; Score 80; DB 4; Length 461;
Best Local Similarity 25.6%; Pred. No. 7;
Matches 43; Conservative 20; Mismatches 67; Indels 38; Gaps 8;

QY 6 VAFAAIVSG-SALAGVVPQGGGNGGNGGNSGDPSTLSIYQVGSANA---ALALQS 61
Db 300 VAALPEVARESGAPSGTAPAGGAA---GGKSPAGLRLARRSPSSATPPAATTYF 356

QY 62 DARKSETTITQSG-YCNGA-----DVQGGADNYDQVLT-----RVVTHEMAHADOWN- 107
Db 357 AARAPPAITEPRSGTGAGDPDRRTAGTCEQGLVVDFAHPAYRLVLTGQIEGDSWNG 416
QY 108 -----AKNSDITVQYCGNNAALVNQTSADSSVMVRQVGFNGNA 146
Db 417 DVFLRIDMAELQGDGFMGIADGNQAVASQVISET-----VGNHA 456

RESULT 14
US-09-252-991A-30710
; Sequence 30710, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30710
; LENGTH: 812
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30710

Query Match 10.3%; Score 80; DB 4; Length 812;
Best Local Similarity 26.7%; Pred. No. 15;
Matches 31; Conservative 10; Mismatches 39; Indels 36; Gaps 5;

QY 51 GSAALALQSDARKSE-----TTTQSGYNGADVGGADNYDQLVTRVWTHMAHAD 104
Db 550 GLNIGVTRSDRRYSERVIVSRSTPSQGLGNLGYGGGASRYQQ-----AD 597
QY 105 -QWAKNSDITVQYG--GN-----NAALVNQTSADSSVMVRQVGF 142
Db 598 LTWRMQNVQGGLYGETGNYTRWADLSGLVMDNAVFASNRINDAFVLVSTKGY 653

RESULT 15
US-09-252-991A-26658
; Sequence 26658, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 26658
; LENGTH: 1034
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-26658

Query Match 10.3%; Score 80; DB 4; Length 1034;
Best Local Similarity 26.7%; Pred. No. 20;
Matches 47; Conservative 19; Mismatches 62; Indels 48; Gaps 10;

QY 14 VSGALAGVVPQGGGNGGNG-----GNSGDPSTLSIYQVGS-ANAALALQSDAR-- 64
Db 748 VDSASASQV-EAGGAGNTGVLGSLSSGGEIFRQASGVYSKGLATGGLIGKAENGGM 806

```

Qy 65 ----KSETTITQSGYNGADVGQADNYDQLV-----TRY----- 95
Db 807 LGNLKASGSVTDQG---GADLGLVGNNSQSAIETAEATGKVS GGSNSRVGG LIGHNLGG 863
Qy 96 -VTHEMAHADQWNAKNSDITVGYGNGNNAALVNTASD--SSVMVRQVG--FGNNA 146
Db 864 SVAHAISRGDVSGGFNS-LVGG L VGHNGGELVNV D ASGRV SAAAASASVGG LVGSNA 918

```

Search completed: August 2, 2004, 14:58:35
Job time : 13 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:54:48 ; Search time 36.8 Seconds
(without alignments)
1287.123 Million cell updates/sec

Title: US-09-543-407-22

Perfect score: 776
Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVQVGFNNATANQY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US05A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US05B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US05C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	518	66.8	151	12	US-09-741-873B-4
2	518	66.8	151	12	US-09-741-873B-4
3	440	56.7	131	12	US-09-741-873B-2
4	440	56.7	131	12	US-09-741-873B-2
5	95	12.2	445	15	US-10-369-493-20638
6	95	12.2	445	15	US-10-369-493-20638
7	94	12.1	297	9	US-09-793-306-146
8	93	12.0	271	16	US-10-437-963-147343
9	93	12.0	369	12	US-10-425-114-56041
10	93	12.0	486	12	US-10-424-599-275468
11	92.5	11.9	251	10	US-10-425-114-57763
12	92.5	11.9	251	10	US-09-880-748-1130
13	91	11.7	1751	9	US-10-293-418-1130
14	91	11.7	1751	9	US-09-841-132-445
15	91	11.7	1751	16	US-10-467-534-45

```

16 89.5 11.5 251 10 US-09-880-748-1122 Sequence 1122, Ap
17 89.5 11.5 251 12 US-10-293-418-1122 Sequence 1122, Ap
18 89 89 193 16 US-10-437-963-148500 Sequence 148500,
19 88.5 11.4 145 16 US-10-437-963-147748 Sequence 147748,
20 88.5 11.4 197 12 US-10-425-114-67750 Sequence 67750, A
21 88 11.3 503 16 US-10-437-963-158876 Sequence 158876,
22 87.5 11.3 154 16 US-10-437-963-162284 Sequence 162284,
23 87.5 11.3 204 12 US-10-424-599-203972 Sequence 203972,
24 87.5 11.3 486 15 US-10-369-493-20619 Sequence 20619, A
25 87.5 11.3 899 16 US-10-437-963-188971 Sequence 188971,
26 87 11.2 892 10 US-09-952-267-5 Sequence 5, Appli
27 86.5 11.1 191 16 US-10-437-963-105413 Sequence 105413,
28 86.5 11.1 238 12 US-10-634-862-42 Sequence 42, Appl
29 86.5 11.1 585 9 US-09-841-132-337 Sequence 337, Ap
30 86.5 11.1 1752 9 US-09-841-132-180 Sequence 180, Ap
31 86 11.1 448 16 US-10-437-963-140685 Sequence 140685,
32 86 11.1 735 12 US-10-425-114-63965 Sequence 63965, A
33 86 11.1 6310 12 US-10-282-122A-67793 Sequence 67793, A
34 85.5 11.0 270 16 US-10-437-963-122263 Sequence 122263,
35 85.5 11.0 580 12 US-10-647-057-4 Sequence 4, Appli
36 85 11.0 354 10 US-09-820-843A-21 Sequence 21, Appl
37 85 11.0 400 16 US-10-437-963-186417 Sequence 186417,
38 84.5 10.9 189 16 US-10-437-963-170736 Sequence 170736,
39 84.5 10.9 1276 16 US-10-437-963-168952 Sequence 168952,
40 84 10.8 242 12 US-10-425-114-61520 Sequence 61520, A
41 84 10.8 253 10 US-09-880-748-2098 Sequence 2098, Ap
42 84 10.8 253 12 US-10-293-418-2098 Sequence 2098, Ap
43 84 10.8 255 10 US-09-880-748-1153 Sequence 1153, Ap
44 84 10.8 255 12 US-10-293-418-1153 Sequence 1153, Ap
45 84 10.8 1448 16 US-10-408-765A-998 Sequence 998, App

```

ALIGNMENTS

RESULT 1

```

US-09-741-873B-4
; Sequence 4, Application US/09741.873B
; Publication No. US20020081722A1
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741.873B
; CURRENT FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-4

```

Query Match 66.8%; Score 518; DB 12; Length 151;
Best Local Similarity 68.9%; Pred. No. 1.2e-44;
Matches 104; Conservative 18; Mismatches 29; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGCGNHNHGGNSGDPSTLSIYQYGSANAALQ 60

```

Db      1  MKLLKVAATAAIVFSSAVAGVVPQYGGGNGHGGGNSGPNSELNIYQYGGNSALALQ 60
QY      61  SDARKSETTITQSGYNGADVGQADNYDQLVTRVVTHEMAHADQNAKNSDITVQYGG 120
Db      61  TDARNSDLTITQGGGNGADVGQSDSSDLTQRFGNSATLDQWNGKNSMTVKQFGG 120
QY      121  NNAALVNOTASDSSVMVROVQFGNNATANQY 151
Db      121  GNGAAVDQIASNSSVNVTVQFGNNATAHQY 151

```

RESULT 2

```

US-09-741-873B-4
; Sequence 4, Application US/09741873B
; Publication No. US2004009695A9
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; PRIOR FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-4

```

```

Query Match          56.7%; Score 440; DB 12; Length 131;
Best Local Similarity 65.6%; Pred. No. 8.3e-37;
Matches 86; Conservative 17; Mismatches 28; Indels 0; Gaps 0;

QY      21  GVVPQWGGGNGHGGGNSGPDSTLSIYQYGSANAALQSDARKSETTITQSGYNGAD 80
Db      1  GVVPYGGGNGHGGGNSGPNSELNIYQYGGNSALALQTDARNSDLTITQGGGNGAD 60

QY      81  VQGSADNYDQLVTRVVTHEMAHADQNAKNSDITVQYGGNNAALVQYASDSSVMVROV 140
Db      61  VQGSDDSSIDLQRFGNSATLDQWNGKNSMTVKQFGGNGAAVDQIASNSSVNVTVQV 120

QY      141  GFGNNATANQY 151
Db      121  GFGNNATAHQY 131

```

RESULT 4

```

US-09-741-873B-2
; Sequence 2, Application US/09741873B
; Publication No. US2004009695A9
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; PRIOR FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-2

```

```

Query Match          66.8%; Score 518; DB 12; Length 151;
Best Local Similarity 68.9%; Pred. No. 1.2e-44;
Matches 104; Conservative 18; Mismatches 29; Indels 0; Gaps 0;

```

```

QY      1  MKLLKVAATAAIVFSSALAGVVPQWGGGNGHGGGNSGPDSTLSIYQYGSANAALQ 60
Db      1  MKLLKVAATAAIVFSSAVAGVVPQYGGGNGHGGGNSGPNSELNIYQYGGNSALALQ 60

QY      61  SDARKSETTITQSGYNGADVGQADNYDQLVTRVVTHEMAHADQNAKNSDITVQYGG 120
Db      61  TDARNSDLTITQGGGNGADVGQSDSSIDLQRFGNSATLDQWNGKNSMTVKQFGG 120

QY      121  NNAALVNOTASDSSVMVROVQFGNNATANQY 151
Db      121  GNGAAVDQIASNSSVNVTVQFGNNATAHQY 151

```

RESULT 3

```

US-09-741-873B-2
; Sequence 2, Application US/09741873B
; Publication No. US20020081722A1
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; PRIOR FILING DATE: 2003-04-04

```

Query Match

56.7%; Score 440; DB 12; Length 131;

Best Local Similarity 65.6%; Pred. No. 8.3e-37;
Matches 86; Conservative 17; Mismatches 28; Indels 0; Gaps 0;
QY 21 GVVPWGNGHNGSGSPDSTLSIYQGSANAALALQSDARKSETTITQSGYNGAD 80
DB 1 GVVPYGGGHHGGGNGSGNSELTNYQGGNSALALQTDARNSDLTITQHGNGAD 60
QY 81 VQGGADNDYDQVTRVVTHEMAHADOWNAKNSDITVQYGGNNAALVNQTSDDSSVMVRQV 140
DB 61 VQGGSDSDSIDLTQRFNGSATLDQWNGKSEMTVKQFGGNGAAMDQGTASNSVNTQV 120
QY 141 GFNNATANQY 151
DB 121 GFNNATAHQY 131
RESULT 5
US-10-369-493-20638
; Sequence 20638, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 20638
; LENGTH: 445
; TYPE: PRT
; ORGANISM: Rhodospseudomonas palustris
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(445)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-20638
Query Match 12.2%; Score 95; DB 15; Length 445;
Best Local Similarity 24.6%; Pred. No. 0.41;
Matches 35; Conservative 22; Mismatches 59; Indels 26; Gaps 5;
QY 30 GNHNGGNS-----GPDSTLSIYQGSANAA-----LALQSDARKSET 68
DB 89 GKQSGAGNSAAIFQEGTGSDELQQTGTSNGAVPSGWNWTPDGVFNKITQDSSNSGSKV 148
QY 69 TITQSGYNGADVGQGDADNDYDQVTRVVTHEMAHADOWN-NAKNSDITVQ---YGGN-NA 123
DB 149 SVIQGKNVVFISIKQNTGNTSVNQIGEWGMAYVVRQIGAAETDASTGNALPTGGNVN 208
QY 124 ALVNQTSDDSSVMVRQVGFNN 145
DB 209 ASITQNSAGLVAVAVQGGNS 230
RESULT 6
US-09-793-306-146
; Sequence 146, Application US/09793306
; Patent No. US20020098200A1
; GENERAL INFORMATION:
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Skelky, Yasir
; APPLICANT: Owendale, Pamela
; APPLICANT: Jen, Shyian
; APPLICANT: Lodes, Michael
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compounds and Methods for Diagnosis and Immunotherapy

; TITLE OF INVENTION: of Tuberculosis
; FILE REFERENCE: 014058-008740US
; CURRENT APPLICATION NUMBER: US/09/793,306
; CURRENT FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: US 60/185,037
; PRIOR FILING DATE: 2000-02-25
; PRIOR APPLICATION NUMBER: US 60/223,828
; PRIOR FILING DATE: 2000-08-08
; NUMBER OF SEQ ID NOS: 164
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 146
; LENGTH: 597
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: mTTC#3-His
US-09-793-306-146
Query Match 12.2%; Score 95; DB 9; Length 597;
Best Local Similarity 27.4%; Pred. No. 0.59;
Matches 34; Conservative 14; Mismatches 50; Indels 26; Gaps 4;
QY 26 WGGGNNHNGGNSGSPDSTLSIYQGSANAALALQSDARKSETTITQSGYNGADYVQGA 85
DB 358 FGNNGNNIGFFNSG-NNNVGFFNSGNN-----FCFGNAGDINTGF 398
QY 86 DNYDQVTRVVTHEMAHADOWNAKNSDITVQYGGNNAALVNQTSDDSSVMVRQVGFNN 145
DB 399 GNAGDNTVGFNGAGFFNMIGNAGNEDMGVNGGSGFNVGVGN--AGNQS-----VGFNA 451
QY 146 ATAN 149
DB 452 GTLN 455
RESULT 7
US-10-437-963-147343
; Sequence 147343, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David X.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 147343
; LENGTH: 271
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_47881C.1.pep
US-10-437-963-147343
Query Match 12.1%; Score 94; DB 16; Length 271;
Best Local Similarity 26.2%; Pred. No. 0.28;
Matches 37; Conservative 19; Mismatches 67; Indels 18; Gaps 6;
QY 14 VSGSALAGVVPWGGG-GNHNGGNSGSPDSTLSIYQGSANAALALQSDARKSETTITQ 72
DB 7 VVGRAAAAYPEVSGDGGSGGGGEGE-----GGDGSVAAVNPEAGSGDGRSS 58
QY 73 SGYNGADYVQGGADNDYDQVTRVVTHEMAHADOWNAKNSDITVQYGGNNAALVNQTSAD 132
DB 59 GGGGGSGGGLGRQRYWNSRLSTERQRLVDR-VFKNSDVVCDVFSG-----VGFIAIS 112

QY 133 SSVVROVQVGFGR--NATANQY 151
DB 113 AARKVRIV-YANDLNPTAVEY 132

RESULT 8

US-10-425-114-56041
; Sequence 56041, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 56041
; LENGTH: 369
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: 701205720_FLI.pep
US-10-425-114-56041

Query Match 12.0%; Score 93; DB 12; Length 369;
Best Local Similarity 26.3%; Pred. No. 0.51;
Matches 36; Conservative 18; Mismatches 63; Indels 20; Gaps 6;

QY 14 VSGSALAGVVPQWGGGNGHNGSGDPSTLSIYQGSANAALQSDARKSETITQS 73
DB 13 VRSSASKG-KPLSAGNNGGNDWDPDD-----GFGSARGG-----ADLRNQSTGTVR 62
QY 74 GYGNAGADVGGADNYDQLVTRVVTTHMAHADQWNAKNSDITVQYGGNNAALVNTQASDS 133
DB 63 GFGGGGVNRSKSTQD-MYTR-----AELEASAANKEDFFARKRAENESRPEGLPPSQG 115
QY 134 SSVVROVQVGFGRNATANQ 150
DB 116 G---KYVFGSGGPAPNQ 129

RESULT 9

US-10-424-599-275468
; Sequence 275468, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 275468
; LENGTH: 486
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_90769C.1.pep
US-10-424-599-275468

Query Match 12.0%; Score 93; DB 12; Length 486;
Best Local Similarity 26.3%; Pred. No. 0.72;

Matches 36; Conservative 18; Mismatches 63; Indels 20; Gaps 6;
QY 14 VSGSALAGVVPQWGGGNGHNGSGDPSTLSIYQGSANAALQSDARKSETITQS 73
DB 130 VRSSASKG-KPLSAGNNGGNDWDPDD-----GFGSARGG-----ADLRNQSTGTVR 179
QY 74 GYGNAGADVGGADNYDQLVTRVVTTHMAHADQWNAKNSDITVQYGGNNAALVNTQASDS 133
DB 180 GFGGGGVNRSKSTQD-MYTR-----AELEASAANKEDFFARKRAENESRPEGLPPSQG 232
QY 134 SSVVROVQVGFGRNATANQ 150
DB 233 G---KYVFGSGGPAPNQ 246

RESULT 10

US-10-425-114-57763
; Sequence 57763, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 57763
; LENGTH: 507
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-GMFLMINSOY049H04_FLI.pep
US-10-425-114-57763

Query Match 12.0%; Score 93; DB 12; Length 507;
Best Local Similarity 26.3%; Pred. No. 0.76;
Matches 36; Conservative 18; Mismatches 63; Indels 20; Gaps 6;

QY 14 VSGSALAGVVPQWGGGNGHNGSGDPSTLSIYQGSANAALQSDARKSETITQS 73
DB 151 VRSSASKG-KPLSAGNNGGNDWDPDD-----GFGSARGG-----ADLRNQSTGTVR 200
QY 74 GYGNAGADVGGADNYDQLVTRVVTTHMAHADQWNAKNSDITVQYGGNNAALVNTQASDS 133
DB 201 GFGGGGVNRSKSTQD-MYTR-----AELEASAANKEDFFARKRAENESRPEGLPPSQG 253
QY 134 SSVVROVQVGFGRNATANQ 150
DB 254 G---KYVFGSGGPAPNQ 267

RESULT 11

US-09-880-748-1130
; Sequence 1130, Application US/09880748
; Publication No. US20030059937A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Antibodies that Immunospecifically Bind Blys
; FILE REFERENCE: PF523
; CURRENT APPLICATION NUMBER: US/09/880,748
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 60/212,210
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: 60/240,816
; PRIOR FILING DATE: 2000-10-17
; PRIOR APPLICATION NUMBER: 60/276,248
; PRIOR FILING DATE: 2001-03-16

```
; PRIOR APPLICATION NUMBER: 60/277,379
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/293,499
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 3239
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1130
; LENGTH: 251
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-880-748-1130
```

```
Query Match 11.9%; Score 92.5; DB 10; Length 251;
Best Local Similarity 27.7%; Pred. No. 0.36;
Matches 33; Conservative 14; Mismatches 51; Indels 21; Gaps 4;
```

```
QY 18 ALAGVVPWG-----GGGNHGGGSSGPDSTLSIYQYGSANAALALQSDARKSE- 67
DB 106 ATTGALDMWGKTLVTVSSGGGGGGGGG-----GSAQAVLTQPSASGTPG 155
QY 68 TTITOSGYNGADVGQADN-YDQLVTRVVTHEMAHADQWNAKNSDITVGOYGGNNAAL 125
DB 156 QRTVMSGSSSNGSNTVNWYQQLPGAAPKLLIYRSQRRSSGVPDRFSGSKSGTSASL 214
```

RESULT 12

```
US-10-293-418-1130
; Sequence 1130, Application US/10293418
; Publication No. US20030223996A1
```

```
; GENERAL INFORMATION:
```

```
; APPLICANT: Ruben et al.
```

```
; TITLE OF INVENTION: Antibodies that Immunospecifically Bind Blys
```

```
; FILE REFERENCE: PF523P2
```

```
; CURRENT APPLICATION NUMBER: US/10/293,418
```

```
; CURRENT FILING DATE: 2002-11-27
```

```
; PRIOR APPLICATION NUMBER: 60/331,469
```

```
; PRIOR FILING DATE: 2001-11-16
```

```
; PRIOR APPLICATION NUMBER: 60/340,817
```

```
; PRIOR FILING DATE: 2001-12-19
```

```
; PRIOR APPLICATION NUMBER: 09/880,748
```

```
; PRIOR FILING DATE: 2001-06-15
```

```
; PRIOR APPLICATION NUMBER: 60/293,499
```

```
; PRIOR FILING DATE: 2001-05-25
```

```
; PRIOR APPLICATION NUMBER: 60/277,379
```

```
; PRIOR FILING DATE: 2001-03-21
```

```
; PRIOR APPLICATION NUMBER: 60/276,248
```

```
; PRIOR FILING DATE: 2001-03-16
```

```
; PRIOR APPLICATION NUMBER: 60/240,816
```

```
; PRIOR FILING DATE: 2000-10-17
```

```
; PRIOR APPLICATION NUMBER: 60/212,210
```

```
; PRIOR FILING DATE: 2000-06-16
```

```
; NUMBER OF SEQ ID NOS: 3247
```

```
; SEQ ID NO 1130
```

```
; LENGTH: 251
```

```
; TYPE: PRT
```

```
; ORGANISM: Homo sapiens
```

```
US-10-293-418-1130
```

```
Query Match 11.9%; Score 92.5; DB 12; Length 251;
Best Local Similarity 27.7%; Pred. No. 0.36;
Matches 33; Conservative 14; Mismatches 51; Indels 21; Gaps 4;
```

```
QY 18 ALAGVVPWG-----GGGNHGGGSSGPDSTLSIYQYGSANAALALQSDARKSE- 67
DB 106 ATTGALDMWGKTLVTVSSGGGGGGGGG-----GSAQAVLTQPSASGTPG 155
QY 68 TTITOSGYNGADVGQADN-YDQLVTRVVTHEMAHADQWNAKNSDITVGOYGGNNAAL 125
DB 156 QRTVMSGSSSNGSNTVNWYQQLPGAAPKLLIYRSQRRSSGVPDRFSGSKSGTSASL 214
```

RESULT 13

```
US-09-841-132-445
```

```
; Sequence 445, Application US/09841132
; Patent No. US20020061848A1
; GENERAL INFORMATION:
; APPLICANT: Bhatia, Ajay
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Probst, Peter
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
; FILE REFERENCE: 210121.469C8
; CURRENT APPLICATION NUMBER: US/09/841,132
; CURRENT FILING DATE: 2001-04-23
; NUMBER OF SEQ ID NOS: 599
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 445
; LENGTH: 1751
; TYPE: PRT
; ORGANISM: Chlamydia trachomatis serovar D
US-09-841-132-445
```

```
Query Match 11.7%; Score 91; DB 9; Length 1751;
Best Local Similarity 33.1%; Pred. No. 5.7;
Matches 40; Conservative 10; Mismatches 59; Indels 12; Gaps 5;
```

```
QY 15 SGSALAGVVPWGCGGN--HNGGNSGSPDSTLSIYQYGSANAALALQSDARKSETTIT- 71
DB 394 NADAWASSSPQSGGATTVNSGDSGSDTSETVPATAKGG-GLYTDKNLSITNITG 452
QY 72 --QSGYNGADVGQADNYDQLVTRVVTHEMAHADQWNAKNSDITVGO-YGGNNAALVNQ 128
DB 453 IIEIANNKATDVGGGA-----YVKGLTTCENSHRLOFLKNSSDKQGGIYGEDNITLSNL 507
QY 129 T 129
DB 508 T 508
```

RESULT 14

```
US-09-841-132-594
```

```
; Sequence 594, Application US/09841132
```

```
; Patent No. US20020061848A1
```

```
; GENERAL INFORMATION:
```

```
; APPLICANT: Bhatia, Ajay
```

```
; APPLICANT: Skeiky, Yasir A.W.
```

```
; APPLICANT: Probst, Peter
```

```
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
```

```
; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
```

```
; FILE REFERENCE: 210121.469C8
```

```
; CURRENT APPLICATION NUMBER: US/09/841,132
```

```
; CURRENT FILING DATE: 2001-04-23
```

```
; NUMBER OF SEQ ID NOS: 599
```

```
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
```

```
; SEQ ID NO 594
```

```
; LENGTH: 1751
```

```
; TYPE: PRT
```

```
; ORGANISM: C. Trachomatis D serovar
```

```
US-09-841-132-594
```

```
Query Match 11.7%; Score 91; DB 9; Length 1751;
Best Local Similarity 33.1%; Pred. No. 5.7;
Matches 40; Conservative 10; Mismatches 59; Indels 12; Gaps 5;
```

```
QY 15 SGSALAGVVPWGCGGN--HNGGNSGSPDSTLSIYQYGSANAALALQSDARKSETTIT- 71
DB 394 NADAWASSSPQSGGATTVNSGDSGSDTSETVPATAKGG-GLYTDKNLSITNITG 452
QY 72 --QSGYNGADVGQADNYDQLVTRVVTHEMAHADQWNAKNSDITVGO-YGGNNAALVNQ 128
DB 453 IIEIANNKATDVGGGA-----YVKGLTTCENSHRLOFLKNSSDKQGGIYGEDNITLSNL 507
QY 129 T 129
DB 508 T 508
```

```
RESULT 15
US-10-467-534-45
; Sequence 45, Application US/10467534
; Publication No. US20040131625A1
; GENERAL INFORMATION:
; APPLICANT: Berthet, Francois-Xavier Jacques
; APPLICANT: Lobet, Yves
; APPLICANT: Poolman, Jan
; APPLICANT: Verlant, Vincent Georges Christian Louis
; TITLE OF INVENTION: Vaccine Composition
; FILE REFERENCE: B45261
; CURRENT APPLICATION NUMBER: US/10/467,534
; CURRENT FILING DATE: 2003-08-08
; PRIOR APPLICATION NUMBER: PCT/EP02/01356
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: GB 0103169.9
; PRIOR FILING DATE: 2001-02-08
; NUMBER OF SEQ ID NOS: 108
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 45
; TYPE: PRT
; LENGTH: 1751
; ORGANISM: Chlamydia trachomatis
US-10-467-534-45

Query Match      11.7%; Score 91; DB 16; Length 1751;
Best Local Similarity 33.1%; Pred. No. 5.7;
Matches 40; Conservative 10; Mismatches 59; Indels 12; Gaps 5;

QY      15  S G S A L A G V P Q W G G G G N - - H N G G G S S G P D S T L S I Y Q Y G S A N A L A L O S D A R K S E T T I T - 71
Db      394  N A D A W A S S P Q S G S G A T T V S N S G D S S G S D S D T S E T V P A T A K G G - G L Y T D K N L S I T N I T G 452

QY      72  - - O S G Y G N G A D V Q G G A D N Y D Q L V T R V T H E M A H A D Q W A K N S D I T V G Q - Y G G N N A A L V N Q 128
Db      453  I I E I A N N K A T D V G G G A - - - - - Y V K G T L T C E N S H R L Q F L K N S S D K Q G G I Y G E D N I T L S N L 507

QY      129  T 129
Db      508  T 508
```

Search completed: August 2, 2004, 15:36:12
Job time : 37.8 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:48:33 ; Search time 167.9 Seconds
(without alignments)
877.809 Million cell updates/sec

Title: US-09-543-407-22

Perfect score: 776

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVGFNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 6019581 seqs, 976053577 residues

Total number of hits satisfying chosen parameters: 6019581

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents AA Main:*

1: /cgn2_6/ptodata/2/paa/PCTUS_COMB.pep.*
2: /cgn2_6/ptodata/2/paa/US06_COMB.pep.*
3: /cgn2_6/ptodata/2/paa/US07_COMB.pep.*
4: /cgn2_6/ptodata/2/paa/US08_COMB.pep.*
5: /cgn2_6/ptodata/2/paa/US081_COMB.pep.*
6: /cgn2_6/ptodata/2/paa/US082_COMB.pep.*
7: /cgn2_6/ptodata/2/paa/US083_COMB.pep.*
8: /cgn2_6/ptodata/2/paa/US084_COMB.pep.*
9: /cgn2_6/ptodata/2/paa/US085_COMB.pep.*
10: /cgn2_6/ptodata/2/paa/US086_COMB.pep.*
11: /cgn2_6/ptodata/2/paa/US087_COMB.pep.*
12: /cgn2_6/ptodata/2/paa/US088_COMB.pep.*
13: /cgn2_6/ptodata/2/paa/US089_COMB.pep.*
14: /cgn2_6/ptodata/2/paa/US090_COMB.pep.*
15: /cgn2_6/ptodata/2/paa/US091_COMB.pep.*
16: /cgn2_6/ptodata/2/paa/US092_COMB.pep.*
17: /cgn2_6/ptodata/2/paa/US093_COMB.pep.*
18: /cgn2_6/ptodata/2/paa/US094_COMB.pep.*
19: /cgn2_6/ptodata/2/paa/US095_COMB.pep.*
20: /cgn2_6/ptodata/2/paa/US096_COMB.pep.*
21: /cgn2_6/ptodata/2/paa/US097A_COMB.pep.*
22: /cgn2_6/ptodata/2/paa/US097B_COMB.pep.*
23: /cgn2_6/ptodata/2/paa/US098_COMB.pep.*
24: /cgn2_6/ptodata/2/paa/US099A_COMB.pep.*
25: /cgn2_6/ptodata/2/paa/US099B_COMB.pep.*
26: /cgn2_6/ptodata/2/paa/US100_COMB.pep.*
27: /cgn2_6/ptodata/2/paa/US101_COMB.pep.*
28: /cgn2_6/ptodata/2/paa/US102_COMB.pep.*
29: /cgn2_6/ptodata/2/paa/US103_COMB.pep.*
30: /cgn2_6/ptodata/2/paa/US104_COMB.pep.*
31: /cgn2_6/ptodata/2/paa/US106_COMB.pep.*
32: /cgn2_6/ptodata/2/paa/US107_COMB.pep.*
33: /cgn2_6/ptodata/2/paa/US60_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
------------	-------	-------	--------	----	-------------

OTHER INFORMATION: Recombinant salmonella enteritidis 3b afgA
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.

1	776	100.0	151	19	US-09-543-407-22	Sequence 22, Appl
2	712	91.8	151	19	US-09-543-407-28	Sequence 28, Appl
3	682	87.9	151	19	US-09-543-407-5	Sequence 5, Appl
4	677	87.2	151	6	US-08-233-642A-57	Sequence 57, Appl
5	657	84.7	151	19	US-09-543-407-30	Sequence 30, Appl
6	639	79.8	151	19	US-09-543-407-20	Sequence 20, Appl
7	609	78.5	151	19	US-09-543-407-24	Sequence 24, Appl
8	604	77.8	151	19	US-09-543-407-12	Sequence 12, Appl
9	602	77.6	151	19	US-09-543-407-14	Sequence 14, Appl
10	601	77.4	151	19	US-09-543-407-26	Sequence 26, Appl
11	600	77.3	151	19	US-09-543-407-18	Sequence 18, Appl
12	595	76.7	131	19	US-09-543-407-31	Sequence 31, Appl
13	585	76.1	131	19	US-09-543-407-16	Sequence 16, Appl
14	521	73.1	151	19	US-09-543-407-7	Sequence 7, Appl
15	518	66.8	151	13	US-08-978-878-4	Sequence 4, Appl
16	518	66.8	151	21	US-09-741-873B-4	Sequence 4, Appl
17	516	66.5	151	33	US-60-352-946-2	Sequence 2, Appl
18	516	66.5	151	33	US-60-444-371-2	Sequence 2, Appl
19	497	64.0	120	6	US-08-233-642A-55	Sequence 55, Appl
20	485	59.9	158	16	US-09-252-691-5834	Sequence 5834, Ap
21	465	59.9	158	16	US-09-252-691C-5834	Sequence 5834, Ap
22	465	59.9	158	30	US-10-417-886-5834	Sequence 5834, Ap
23	463	59.7	109	19	US-09-543-407-34	Sequence 34, Appl
24	440	56.7	131	13	US-08-978-878-2	Sequence 2, Appl
25	440	56.7	131	21	US-09-741-873B-2	Sequence 2, Appl
26	336	43.3	109	19	US-09-543-407-35	Sequence 35, Appl
27	256	33.0	68	19	US-09-543-407-37	Sequence 37, Appl
28	215	27.7	48	19	US-09-543-407-39	Sequence 39, Appl
29	159.5	20.6	70	19	US-09-543-407-32	Sequence 32, Appl
30	104.5	13.5	186	16	US-09-252-691-5833	Sequence 5833, Ap
31	104.5	13.5	186	16	US-09-252-691C-5833	Sequence 5833, Ap
32	104.5	13.5	186	30	US-10-417-886-5833	Sequence 5833, Ap
33	97.5	12.6	520	1	PCT-US02-18256-21	Sequence 21, Appl
34	96	12.4	2308	33	US-60-161-932-688	Sequence 688, App
35	96	12.4	2309	20	US-09-614-150-25488	Sequence 25488, A
36	96	12.4	2309	20	US-09-614-150A-25488	Sequence 25488, A
37	96	12.4	2309	33	US-60-191-637-25607	Sequence 25607, A
38	96	12.4	2309	33	US-60-191-681-20216	Sequence 20216, A
39	95.5	12.3	252	1	PCT-US03-13414-24	Sequence 24, Appl
40	95	12.2	445	29	US-10-369-493-20638	Sequence 20638, A
41	95	12.2	445	33	US-60-360-039-20638	Sequence 20638, A
42	95	12.2	597	1	PCT-US01-05992-146	Sequence 146, App
43	95	12.2	597	22	US-09-793-305-146	Sequence 146, App
44	94	12.1	271	30	US-10-437-963-147343	Sequence 147343,
45	93.5	12.0	145	21	US-09-739-449-8854	Sequence 8854, Ap

ALIGNMENTS

RESULT 1
US-09-543-407-22
; Sequence 22, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 22
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.

US-09-543-407-22

Query Match 100.0%; Score 776; DB 19; Length 151;
Best Local Similarity 100.0%; Pred. No. 1.2e-74;
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60

Qy 61 SDARKSETTITQSGYNGADVGGADNYDQLVTRVVTHEMAHADQWNAKNSDITVQYGG 120
Db 61 SDARKSETTITQSGYNGADVGGADNYDQLVTRVVTHEMAHADQWNAKNSDITVQYGG 120

Qy 121 NNAALVNQASDSSVMVRQVGFNNATANQY 151
Db 121 NNAALVNQASDSSVMVRQVGFNNATANQY 151

RESULT 2

US-09-543-407-28
Sequence 28, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
FILE REFERENCE: 920043.406
PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
CURRENT APPLICATION NUMBER: US/09/543,407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 28
LENGTH: 151
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.

US-09-543-407-28

Query Match 91.8%; Score 712; DB 19; Length 151;
Best Local Similarity 91.1%; Pred. No. 9.1e-68;
Matches 144; Conservative 0; Mismatches 0; Indels 14; Gaps 2;

Qy 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60

Qy 61 SDARKSETTITQSGYNGADVGGADNYDQLVTRVVTHEMAHADQWNAKNSDI 113
Db 61 SDARKSETTITQSGYNGADVGGADNYDQLVTRVVTHEMAHADQWNAKNSDI 113

Qy 114 TVGQYGGNNAALVNQASDSSVMVRQVGFNNATANQY 151
Db 114 TVGQYGGNNAALVNQASDSSVMVRQVGFNNATANQY 151

RESULT 3

US-09-543-407-5
Sequence 5, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
FILE REFERENCE: 920043.406
PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
CURRENT APPLICATION NUMBER: US/09/543,407

CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5
LENGTH: 151
TYPE: PRT
ORGANISM: Salmonella enteritidis
US-09-543-407-5

Query Match 87.9%; Score 682; DB 19; Length 151;
Best Local Similarity 90.7%; Pred. No. 1.5e-64;
Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

Qy 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60

Qy 61 SDARKSETTITQSGYNGADVGGADNYDQLVTRVVTHEMAHADQWNAKNSDITVQYGG 120
Db 61 SDARKSETTITQSGYNGADVGGADNYDQLVTRVVTHEMAHADQWNAKNSDITVQYGG 120

Qy 121 NNAALVNQASDSSVMVRQVGFNNATANQY 151
Db 121 NNAALVNQASDSSVMVRQVGFNNATANQY 151

RESULT 4

US-08-233-642A-57
Sequence 57, Application US/08233642A
GENERAL INFORMATION:
APPLICANT: Kay, William W.
APPLICANT: Collinson, S. Karen
APPLICANT: Clouthier, Sharon C.
APPLICANT: Doran, James L.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
TITLE OF INVENTION: BASED VACCINES
NUMBER OF SEQUENCES: 58
NUMBER OF SEQUENCES: -
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed and Berry
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: U.S.A.
ZIP: 98104-7092
COMPUTER READABLE FORM: disk
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233,642A
FILING DATE: 26-APR-1994
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: King, Joshua
REGISTRATION NUMBER: 35,570
REFERENCE/DOCKET NUMBER: 920043.403C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEEDANBERY
INFORMATION FOR SEQ ID NO: 57
SEQUENCE CHARACTERISTICS:
LENGTH: 151 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-233-642A-57

Query Match 87.2%; Score 677; DB 6; Length 151;
Best Local Similarity 90.1%; Pred. No. 5.2e-64;
Matches 136; Conservative 2; Mismatches 13; Indels 0; Gaps 0;


```
QY 1 MKLLKVAAPAAIVVSGSALAGVVPQGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
Db 1 MKLLKVAAPAAIVVSGSALAGVVPQGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADYVGGADNYDQLVTRVVTHEMAHADQNAKNSDITVQYGG 120
Db 61 SDARKSETTITQSGYNGADYVGGADNYDQLVTRVVTHEMAHADQNAKNSDITVQYGG 120
QY 121 NNAALVNQTSASDSSVMVROVGFQGNATANOY 151
Db 121 NNPALVNQTSASDSSVMVROVGFQGNATANOY 151

RESULT 5
US-09-543-407-30
; Sequence 30, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-30
Query Match 84.7%; Score 657; DB 19; Length 151;
Best Local Similarity 81.9%; Pred. No. 7.4e-62;
Matches 136; Conservative 0; Mismatches 0; Indels 30; Gaps 2;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
Db 1 MKLLKVAAPAAIVVSGSALAGVVPQGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADYVGGADNYDQLVTRVVTHEMAHADQNAKNSDITVQYGG 105
Db 61 SDARKSETTITQSGYNGADYVGGADNYDQLVTRVVTHEMAHADQNAKNSDITVQYGG 118
QY 106 WNAKNSDITVGGYGGNNAALVNQTSASDSSVMVROVGFQGNATANOY 151
Db 119 -----GGNNAALVNQTSASDSSVMVROVGFQGNATANOY 151

RESULT 6
US-09-543-407-20
; Sequence 20, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-20
Query Match 78.5%; Score 609; DB 19; Length 151;
Best Local Similarity 73.6%; Pred. No. 1.1e-56;
Matches 128; Conservative 0; Mismatches 0; Indels 46; Gaps 2;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
Db 1 MKLLKVAAPAAIVVSGSALAGVVPQGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADYVGGADNYDQLVTRVVTHEMAHADQNAKNSDITVQYGG 97
Db 61 SDARKSETTITQSGYNGADYVGGADNYDQLVTRVVTHEMAHADQNAKNSDITVQYGG 120
QY 98 HEVAHADQNAKNSDITVGGYGGNNAALVNQTSASDSSVMVROVGFQGNATANOY 151
Db 121 HEMAHA-----NQTSASDSSVMVROVGFQGNATANOY 151

RESULT 8
US-09-543-407-12
; Sequence 12, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
```

```
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-20
Query Match 79.8%; Score 619; DB 19; Length 151;
Best Local Similarity 74.6%; Pred. No. 8.9e-58;
Matches 129; Conservative 0; Mismatches 0; Indels 44; Gaps 2;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
Db 1 MKLLKVAAPAAIVVSGSALAGVVPQGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADYVGGADNYDQLVTRVVTHEMAHADQNAKNSDITVQYGG 103
Db 61 SDARK-----YDQLVTRVVTHEMAHADQNAKNSDITVQYGG 98
QY 104 -----DOWNAKNSDITVGGYGGNNAALVNQTSASDSSVMVROVGFQGNATANOY 151
Db 99 NNATIDOWNAKNSDITVGGYGGNNAALVNQTSASDSSVMVROVGFQGNATANOY 151

RESULT 7
US-09-543-407-24
; Sequence 24, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-24
Query Match 78.5%; Score 609; DB 19; Length 151;
Best Local Similarity 73.6%; Pred. No. 1.1e-56;
Matches 128; Conservative 0; Mismatches 0; Indels 46; Gaps 2;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
Db 1 MKLLKVAAPAAIVVSGSALAGVVPQGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADYVGGADNYDQLVTRVVTHEMAHADQNAKNSDITVQYGG 97
Db 61 SDARKSETTITQSGYNGADYVGGADNYDQLVTRVVTHEMAHADQNAKNSDITVQYGG 120
QY 98 HEVAHADQNAKNSDITVGGYGGNNAALVNQTSASDSSVMVROVGFQGNATANOY 151
Db 121 HEMAHA-----NQTSASDSSVMVROVGFQGNATANOY 151

RESULT 8
US-09-543-407-12
; Sequence 12, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
```



```
/ EARLIER APPLICATION NUMBER: US 07/347,189
/ EARLIER FILING DATE: 1989-05-04
/ EARLIER APPLICATION NUMBER: US 07/789,437
/ EARLIER FILING DATE: 1991-11-06
/ EARLIER APPLICATION NUMBER: US 07/970,846
/ EARLIER FILING DATE: 1992-11-03
/ EARLIER APPLICATION NUMBER: US 08/187,865
/ EARLIER FILING DATE: 1994-01-28
/ EARLIER APPLICATION NUMBER: US 08/318,519
/ EARLIER FILING DATE: 1994-10-05
/ EARLIER APPLICATION NUMBER: US 08/495,959
/ EARLIER FILING DATE: 1995-06-28
/ NUMBER OF SEQ ID NOS: 10
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 4
/ LENGTH: 151
/ TYPE: PRT
/ ORGANISM: Escherichia coli
US-08-978-878-4

Query Match      56.8%; Score 518; DB 13; Length 151;
Best Local Similarity 68.9%; Pred. No. 6.3e-47;
Matches 104; Conservative 18; Mismatches 29; Indels 0; Gaps 0;

QY      1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHGGNSGDPDTLSIYQYGSANAALALQ 60
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      1 MKLLKVAALAAIVFSGSAVAGVVPQYGGGNGHGGNSGPNSELNIYQYGGNSALALQ 60
      :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||

QY      61 SDARKEETITQSGYNGADVCGGADNDYDQLVTRVVTHEMAHADOWNAKNSDITVGYGG 120
      :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
Db      61 TDARNSDLTITQGGGNGADVCGGSDSSIDLTRQFGNSATLDWNGKNSMTVKQFGG 120
      :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||

QY      121 NNAALVNOTASDSSVMVROVGFNNATANOY 151
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      121 GNGAAVDQIASNSSVNVTVQVFGNNATAHOY 151
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
```

Search completed: August 2, 2004, 15:26:45
Job time : 168.9 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:49:38 ; Search time 17.8 seconds
(without alignments)
888.146 Million cell updates/sec

Title: US-09-543-407-22
Perfect score: 776
Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVQVGGNNATANQY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 601315 seqs, 104695340 residues

Total number of hits satisfying chosen parameters: 601315

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents AA, New:*
1: /cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pep.*
2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep.*
3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep.*
4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep.*
5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep.*
6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep.*
7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep.*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	53.8	66.8	151	5	US-09-741-873C-4
2	44.0	56.7	131	5	US-09-741-873C-2
3	97.5	12.6	520	6	US-10-479-638-21
4	92	11.9	841	7	US-60-565-632-7906
5	92	11.9	841	7	US-60-579-062-7906
6	91	11.7	1751	6	US-10-498-327-1115
7	91	11.7	1751	6	US-10-872-155-445
8	91	11.7	1751	6	US-10-872-155-594
9	90	11.6	1010	1	PCT-US04-21432-402
10	89.5	11.5	388	5	US-09-248-796A-17306
11	88.5	11.4	193	6	US-10-425-115-254240
12	87	11.2	892	5	US-09-952-267B-5
13	87	11.2	892	6	US-10-872-769-5
14	87	11.2	892	6	US-10-872-769-5
15	86.5	11.1	125	6	US-10-425-115-33419
16	86.5	11.1	585	6	US-10-872-153-337
17	86.5	11.1	1752	6	US-10-872-153-180
18	86	11.1	201	6	US-10-425-115-309662
19	86	11.1	244	6	US-10-854-439-84
20	85	11.0	197	6	US-10-425-115-304391
21	85	11.0	215	5	US-09-248-796A-20202
22	85	11.0	239	6	US-10-425-115-343835
23	85	11.0	293	6	US-10-425-115-312468
24	85	11.0	573	7	US-60-565-632-7907
25	85	11.0	573	7	US-60-579-062-7907
26	85	11.0	688	7	US-60-581-351-11388

27	84.5	10.9	282	6	US-10-501-282-2420	Sequence 2420, Ap
28	84.5	10.9	303	6	US-10-501-282-2422	Sequence 2422, Ap
29	84	10.8	256	6	US-10-425-115-301334	Sequence 301334, A
30	84	10.8	619	1	PCT-US04-10229-38	Sequence 38, Appl
31	84	10.8	753	6	US-10-170-205E-35514	Sequence 35514, A
32	84	10.8	753	6	US-10-170-205E-35515	Sequence 35515, A
33	84	10.8	1627	6	US-10-170-205E-16659	Sequence 16659, A
34	84	10.8	1905	1	PCT-US04-09388-9	Sequence 9, Appl
35	83	10.7	586	1	PCT-US03-24982A-317	Sequence 317, Appl
36	82.5	10.6	246	6	US-10-854-439-511	Sequence 511, App
37	82.5	10.6	687	6	US-10-425-115-278239	Sequence 278239, A
38	82	10.6	374	1	PCT-US04-11210-37	Sequence 37, Appl
39	81.5	10.5	508	6	US-10-425-115-285216	Sequence 285216, A
40	81.5	10.5	532	6	US-10-425-115-285214	Sequence 285214, A
41	81.5	10.5	669	6	US-10-489-425-18	Sequence 18, Appl
42	81	10.4	412	7	US-60-565-632-7905	Sequence 7905, Ap
43	81	10.4	412	7	US-60-579-062-7905	Sequence 7905, Ap
44	81	10.4	1358	6	US-10-778-804-11	Sequence 11, Appl
45	80.5	10.4	234	6	US-10-767-701-45603	Sequence 45603, A

ALIGNMENTS

RESULT 1
US-09-741-873C-4
; Sequence 4, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873C-4

Query Match 66.8% Score 518; DB 5; Length 151;
Best Local Similarity 68.9%; Pred.No. 3.4e-37;
Matches 104; Conservative 18; Mismatches 29; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNNHGGNSGPDSTLSTIYQVGSNAALALQ 60
1 MKLLKVAIAAIVFSGSAVAGVVPQYGGGNGHGGNNSGPNSELNTIYQVGGNSALALQ 60
DB 1
QY 61 SDARKSETTITQSYGNGADVGQGDNDYDQIVRVVTHEVAHADQWNAKNSDITVGYGG 120
61 TDAENSLITITQHGNGADVGQGDSSIDLQGRFGNSATLDQWGNKSEMTVKVFGG 120
QY 121 NNAALVNOTASDSSVMVQVQFGNNATANQY 151
DB 121 GNGRAVDOTASNSSVNVTVQVFGNNATAHQY 151

RESULT 2
 US-09-741-873C-2
 ; Sequence 2, Application US/09741873C
 ; GENERAL INFORMATION:
 ; APPLICANT: Norkmark, Staffan
 ; APPLICANT: Olsen, Arne
 ; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
 ; FILE REFERENCE: 012889-084
 ; CURRENT APPLICATION NUMBER: US/09/741,873C
 ; CURRENT FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: SE 8801723-1
 ; PRIOR FILING DATE: 1998-05-06
 ; PRIOR APPLICATION NUMBER: US 08/978,878
 ; PRIOR FILING DATE: 1997-11-26
 ; PRIOR APPLICATION NUMBER: US 07/347,189
 ; PRIOR FILING DATE: 1989-05-04
 ; PRIOR APPLICATION NUMBER: US 07/789,437
 ; PRIOR FILING DATE: 1991-11-06
 ; PRIOR APPLICATION NUMBER: US 07/970,846
 ; PRIOR FILING DATE: 1992-11-03
 ; PRIOR APPLICATION NUMBER: US 08/187,865
 ; PRIOR FILING DATE: 1994-01-28
 ; PRIOR APPLICATION NUMBER: US 08/318,519
 ; PRIOR FILING DATE: 1994-10-05
 ; NUMBER OF SEQ ID NOS: 11
 ; SOFTWARE: Patentin version 3.0
 ; SEQ ID NO 2
 ; LENGTH: 131
 ; TYPE: PRT
 ; ORGANISM: Escherichia coli
 US-09-741-873C-2

Query Match 56.7%; Score 440; DB 5; Length 131;
 Best Local Similarity 65.6%; Pred. No. 1.4e-30;
 Matches 86; Conservative 17; Mismatches 28; Indels 0; Gaps 0;
 QY 21 GVVPQGGGNNHNGGNSGDPSTLSIYQYGSANAALALQSDARKSETTITQSGYNGAD 80
 DB 1 GVVPQYGGGNNHNGGNSGPNSELNTYQYGGNSALALQTDARNSLTITQHGGNGAD 60
 QY 81 VGGADNYDQLVTRVTHEMAHADOWNAKNSDITVQYGGNNAALVNOTASDSSVMYQV 140
 DB 61 VGGSDSSIDLTFQFGNSATLDQNGKNSEMTVQFGGNGAAVDDQTASNSSVNVTVQ 120
 QY 141 GFGNNATANQY 151
 DB 121 GFGNNATAHQY 131

RESULT 3
 US-10-479-638-21
 ; Sequence 21, Application US/10479638
 ; GENERAL INFORMATION:
 ; APPLICANT: Don A. Roth
 ; APPLICANT: Randolph V. Lewis
 ; APPLICANT: The University of Wyoming
 ; TITLE OF INVENTION: Expression of Spider Silk Proteins in Higher Plants
 ; FILE REFERENCE: WYO.02-0004US
 ; CURRENT APPLICATION NUMBER: US/10/479,638
 ; CURRENT FILING DATE: 2003-12-03
 ; PRIOR APPLICATION NUMBER: PCT/US02/18256
 ; PRIOR FILING DATE: 2002-06-06
 ; PRIOR APPLICATION NUMBER: 60/296,184
 ; PRIOR FILING DATE: 2001-06-06
 ; NUMBER OF SEQ ID NOS: 54
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 21
 ; LENGTH: 520
 ; TYPE: PRT
 ; ORGANISM: Argiope trifasciata
 US-10-479-638-21

Query Match 12.6%; Score 97.5; DB 6; Length 520;

Best Local Similarity 22.9%; Pred. No. 1.7;
 Matches 33; Conservative 24; Mismatches 62; Indels 25; Gaps 4;
 QY 15 SGSALAGVVPQWGGGNNHNGGNSGDPSTLSIYQYGSANAALALQSDARKSET-TITQS 73
 DB 347 AGAGAAAASAGAGAGYGGYGVAGGSS-----ISYCATSSSTSSSTASSRSRGIVTSG 402
 QY 74 GYNGADV-----QGADNYDQLVTRVTHEMAHADOWNAKNSDITVQYGGNNAALVN 127
 DB 403 GYGAGAAAAGAGAGAGAGSYGSGISRLSAAE-----AVNRVSSNIGAVAS 448
 QY 128 QTASDSSVMYQVQFGNNATANQY 151
 DB 449 GGASALPGVISNIFSGVSSSAGSY 472
 RESULT 4
 US-60-565-632-7906
 ; Sequence 7906, Application US/60565632
 ; GENERAL INFORMATION:
 ; APPLICANT: Monsanto Technology, LLC
 ; APPLICANT: Baum, James A
 ; APPLICANT: Kovalic, David K.
 ; APPLICANT: Larosa, Thomas J.
 ; APPLICANT: Lu, Maolong
 ; APPLICANT: Munyikwa, Tichifa R. I.
 ; APPLICANT: Roberts, James K.
 ; APPLICANT: Wu, Wei
 ; APPLICANT: Zhang, Bei
 ; TITLE OF INVENTION: Methods for Genetic Control of Insect Infestations in Plants and
 ; FILE REFERENCE: 38-21(53403)B
 ; CURRENT APPLICATION NUMBER: US/60/565,632
 ; CURRENT FILING DATE: 2004-04-27
 ; NUMBER OF SEQ ID NOS: 15449
 ; SOFTWARE: Patentin version 3.2
 ; SEQ ID NO 7906
 ; LENGTH: 841
 ; TYPE: PRT
 ; ORGANISM: Diabrotica virgifera
 ; NAME/KEY: misc feature
 ; LOCATION: (810)..(810)
 ; OTHER INFORMATION: Xaa can be any naturally occurring amino acid
 US-60-565-632-7906

Query Match 11.9%; Score 92; DB 7; Length 841;
 Best Local Similarity 28.3%; Pred. No. 8.9; Indels 16; Gaps 5;
 Matches 36; Conservative 16; Mismatches 59;
 QY 30 GNHNGGG--NSSGPDSTLSIYQYGSANAALALQSDARKSETTITQSG-----YNGADV 81
 DB 657 GNENSGAAENTGNADSQNDAGQ--GSANAA-----NADNANTDAQNGAQGNENGSAAEI 710
 QY 82 GQADNYDQLVTRVTHEMAHADOWNAKNSDITVQYGGNNAALVNOTASDSSVMYQV 141
 DB 711 NGAGDQNE--NGAAAAASGNADQNNNNNSNDQNNNGNSVEENSKEDSGNESQEN 768
 QY 142 FGNNTATA 148
 DB 769 KDNNDTS 775

RESULT 5
 US-60-579-062-7906
 ; Sequence 7906, Application US/60579062
 ; GENERAL INFORMATION:
 ; APPLICANT: Baum, James A
 ; APPLICANT: Kovalic, David K
 ; APPLICANT: Larosa, Thomas J
 ; APPLICANT: Lu, Maolong
 ; APPLICANT: Munyikwa, Tichifa R. I.
 ; APPLICANT: Roberts, James K

```

; APPLICANT: Wu, Wei
; APPLICANT: Zhang, Bei
; TITLE OF INVENTION: Methods for Genetic Control of Insect Infestations in Plants and
; FILE OF INVENTION: Compositions thereof
; FILE REFERENCE: 38-21 (53403) C
; CURRENT APPLICATION NUMBER: US/60/579,062
; CURRENT FILING DATE: 2004-06-11
; NUMBER OF SEQ ID NOS: 41445
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 7506
; LENGTH: 841
; TYPE: PRT
; ORGANISM: Diabrotica virgifera
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (810)..(810)
; OTHER INFORMATION: Xaa can be any naturally occurring amino acid
US-60-579-062-7906

```

Query Match	11.9%;	Score 92;	DB 7;	Length 841;
Best Local Similarity	28.3%;	Pred. No. 8.9;		
Matches	36;	Conservative 16;	Mismatches 59;	Indels 16; Gaps 5;

Qy	30	GNHCGG--NSSGPDSTLSIYQYGSNAALALOSDARKSETITQSG-----YNGADV	81
Db	657	GNENGSAAETGNADSQNDAGQ-GSAAW-----NADNNATDAONGADQNGENGSAAEI	710
Qy	82	GCGADNYQLVTRVVTTHMAFAHDOWNAKNSDITVQYGGNNAALYNOTASDSSVMVRQVG	141
Db	711	NGGADQNE--NGAAASSGNVADNQNNNSNNDNSGNQNGNSVTEENSKEDSGNESQEN	768
Qy	142	FGNNATA	148
Db	769	KDNDTS	775

```

RESULT 6
US-10-498-327-115
; Sequence 115, Application US/10498327
; GENERAL INFORMATION:
; APPLICANT: Grandi, Guido
; APPLICANT: Ratti, Giulio
; TITLE OF INVENTION: Immunisation Against Chlamydia Trachomatis
; FILE REFERENCE: 002441.00085
; CURRENT APPLICATION NUMBER: US/10/498,327
; CURRENT FILING DATE: 2004-06-10
; PRIOR APPLICATION NUMBER: PCT/IB02/05761
; PRIOR FILING DATE: 2002-12-12
; NUMBER OF SEQ ID NOS: 262
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 115
; LENGTH: 1751
; TYPE: PRT
; ORGANISM: Chlamydia trachomatis
US-10-498-327-115

```

Query Match	11.7%;	Score 91;	DB 6;	Length 1751;
Best Local Similarity	33.1%;	Pred. No. 25;		
Matches	40;	Conservative 10;	Mismatches 59;	Indels 12; Gaps 5
Qy	15	SGSALAGVVPQWGGGN--HNGGNSGPDSTLSIYQYGSANAALALQSDARKSETTIT-	71	
Dd	394	NADAWASSPQSGSGATTVNSGSSSGSDSTETVPATAKGG-GLYTDKNLSITNITG	452	
Qy	72	--QSGYGCNADYGGQADNVQDLIVRVVTHEMAHADQWNKNSDITYGQ-YGNNNAALVNQ	128	
Dd	453	IIEIANNKATDVGGGA-----YVKGTLTCENSHRLQFLKNSSDKQGGIYGEDNITLSNL	507	
Qy	129	T	129	
Dd	508	T	508	

RESULT 7
US-10-872-155-445
; Sequence 445, Application US/10872155
; GENERAL INFORMATION:
; APPLICANT: Bhatia, Ajay
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Probst, Peter
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
; FILE REFERENCE: 210121.469C10
; CURRENT APPLICATION NUMBER: US/10/872,155
; CURRENT FILING DATE: 2004-06-18
; PRIOR APPLICATION NUMBER: 09/620,412
; PRIOR FILING DATE: 2000-07-20
; PRIOR APPLICATION NUMBER: 09/598,419
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/556,877
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: 09/454,684
; PRIOR FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: 09/426,571
; PRIOR FILING DATE: 1999-10-22
; PRIOR APPLICATION NUMBER: 09/410,568
; PRIOR FILING DATE: 1999-10-01
; PRIOR APPLICATION NUMBER: 09/288,594
; PRIOR FILING DATE: 1999-04-08
; PRIOR APPLICATION NUMBER: 09/208,277
; PRIOR FILING DATE: 1998-12-08
; NUMBER OF SEQ ID NOS: 599
; SOFTWARE: PastSeq for Windows Version 3.0/4.0
; SEQ ID NO 445
; LENGTH: 1751
; TYPE: PR1
; ORGANISM: Chlamydia trachomatis serovar D
US-10-872-155-445

```

Query Match      11.7%; Score 91; DB 6; Length 1751;
Best Local Similarity 33.1%; Pred.No. 25;
Matches 40; Conservative 10; Mismatches 59; Indels 12; Gaps 5;

Qy      15  S G S A L A G V P Q W G G G N - - H N G G N S S G P D S T L S I Y Q Y G S A N A A L A I Q S D A R K S E T T I T - 71
Db      394  N A D A W S S S F Q S G S G A T T V S N G D S S G S D S D T S E T V P A T A K G G - G L Y T D K N L S I T N I T G 452

Qy      72  - - Q S G Y G N G A D V G G G A D N Y Q L Y T R V V T H E M A H A D Q W A K N S D I T V C Q - Y G G N N A A L V N Q 128
Db      453  I I E I A N N A K A T D V G G A - - - - - Y V K G T L T C E N S H R L Q F L K N S D K Q G G G I Y G E D N I T L S N L 507

Qy      129  T 129
Db      508  T 508

RESULT 8
US-10-872-155-594
; Sequence 594, Application US/10872155
; GENERAL INFORMATION:
; APPLICANT: Bhatia, Ajay
; APPLICANT: Steiky, Yasir A.W.
; APPLICANT: Probst, Peter
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
; FILE REFERENCE: 210121.469C10
; CURRENT APPLICATION NUMBER: US/10/872,155
; CURRENT FILING DATE: 2004-06-18
; PRIOR APPLICATION NUMBER: 09/620,412
; PRIOR FILING DATE: 2000-07-20
; PRIOR APPLICATION NUMBER: 09/598,419
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/556,877
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: 09/454,684
; PRIOR FILING DATE: 1999-12-03

```

```
; PRIOR APPLICATION NUMBER: 09/426,571
; PRIOR FILING DATE: 1999-10-22
; PRIOR APPLICATION NUMBER: 09/410,568
; PRIOR FILING DATE: 1999-10-01
; PRIOR APPLICATION NUMBER: 09/288,594
; PRIOR FILING DATE: 1999-04-08
; PRIOR APPLICATION NUMBER: 09/208,277
; PRIOR FILING DATE: 1998-12-08
; NUMBER OF SEQ ID NOS: 599
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 594
; LENGTH: 1751
; TYPE: PRT
; ORGANISM: C. Trachomatis D serovar
US-10-872-155-594

Query Match      11.7%; Score 91; DB 6; Length 1751;
Best Local Similarity 33.1%; Pred. No. 25;
Matches 40; Conservative 10; Mismatches 59; Indels 12; Gaps 5;

QY 15 SCSALAGVVPQGGG--HNGGGSSGDPSTLSIYQGSANAALQSDARKSETTIT- 71
Db 394 NADAWASSPQSGGATTVNSGDSGSDSTSTVPATKGG-GLYTDKNLSITNITG 452
QY 72 --QSGYGNGADYQGQADNYDQLVTRVVTHEMAHADQNAKNSDITVQY-YGNNAAALVNQ 128
Db 453 IIEIANNKATDVGGGA-----YVKGTLICENSHRLQFLKNSSDKQGGIYGEDNITLSNL 507
QY 129 T 129
Db 508 T 508

RESULT 9
PCT-US04-21492-402
; Sequence 402, Application PC/TUS0421492
; GENERAL INFORMATION:
; APPLICANT: Diversa Corporation
; APPLICANT: Steer, Brian
; APPLICANT: Callen, Walter
; APPLICANT: Healey, Shaun
; APPLICANT: Pulliam, Derrick
; TITLE OF INVENTION: GLUCANASES, NUCLEIC ACIDS ENCODING THEM AND METHODS FOR MAKING AN
; FILE REFERENCE: 564462009540
; CURRENT APPLICATION NUMBER: PCT/US04/21492
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: 60/484,725
; PRIOR FILING DATE: 2003-07-02
; NUMBER OF SEQ ID NOS: 518
; SOFTWARE: Patent In version 3.1
; SEQ ID NO 402
; LENGTH: 1010
; TYPE: PRT
; ORGANISM: Bacteria
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)...(30)
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (393)...(428)
; OTHER INFORMATION: Carbohydrate binding domain
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (39)...(300)
; OTHER INFORMATION: Cellulase (glycosyl hydrolase family 5)
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (610)...(959)
; OTHER INFORMATION: Glycosyl hydrolases family 6
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (493)...(521)
```

```
; OTHER INFORMATION: Cellulose or protein binding domain
PCT-US04-21492-402

Query Match      11.6%; Score 90; DB 1; Length 1010;
Best Local Similarity 22.4%; Pred. No. 16;
Matches 32; Conservative 27; Mismatches 56; Indels 28; Gaps 5;

QY 12 IVVSGSALAGVVPQGW---GGNHNGGSSGDPSTLSIYQGSANAALQSDARKSET 68
Db 316 LTASGTLVKNIVNWTGTTGNGSSSSSSSSSSSS-----SSSSSSSSSSSSSSSS 367
QY 69 TITQSGYNGCA-----DVGQGANVDQLVTRVVTHEMAHADQW---NAKNSD-- 112
Db 368 SSGSTGGNCAGVNVVPTNWTARDWSGAINHANAGDQWVQNSLYRANWYTNVPSGSDAS 427
QY 113 -ITVGYGNNAAALVNQTSDDS 134
Db 428 WTSLGACGGNGSTTSSSSSSSS 450

RESULT 10
US-09-248-796A-17306
; Sequence 17306, Application US/09248796A
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 17306
; LENGTH: 388
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-17306

Query Match      11.5%; Score 89.5; DB 5; Length 388;
Best Local Similarity 22.1%; Pred. No. 6;
Matches 33; Conservative 17; Mismatches 44; Indels 55; Gaps 4;

QY 30 GNHNGGSSGDPSTLSIYQGSANAALQSDARKSETTITQSGYNGADYQGADNYD 89
Db 132 GNQVGSNSYSPDT-----YGSAIGTLGQEKTPAVTGIHSGIGAAA-----YP 177
QY 90 QLVTRVVTHEMAHADQNAKNSDITVGYGNNAAALVNQTS----- 131
Db 178 ELT-----NAGNTGLAKGTAPASTATYGESPSADYKSGATGVPATYLN 223
QY 132 -----DSSVMVRQVGFNNATANQY 151
Db 224 TSGAPTGSINTAGVGGAGFGDNTSSY 252

RESULT 11
US-10-425-115-254240
; Sequence 254240, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 254240
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:39:53 ; Search time 9.4 Seconds
(without alignments)
1545.204 Million cell updates/sec

Title: US-09-543-407-22

Perfect score: 776

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVGFNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78:**

1: Pir1.*

2: Pir2.*

3: Pir3.*

4: Pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	682	87.9	151	2	JC6039	fimbrin protein ag
2	682	87.9	151	2	A10635	major curlin chain
3	521	67.1	151	2	S70788	curlin protein csq
4	499.5	64.4	152	2	D90806	curlin major subun
5	499.5	64.4	152	2	H85665	hypothetical prote
6	109	14.0	1748	2	S42136	cnjB protein - Tet
7	109	14.0	2174	2	E95965	hypothetical glyci
8	102	13.1	586	2	T26667	hypothetical prote
9	98	12.6	615	2	E70663	probable pPE prote
10	95	12.2	590	2	E70946	probable pPE prote
11	94	12.1	407	2	T21956	hypothetical prote
12	93.5	12.0	145	2	AD3143	conserved hypotet
13	93.5	12.0	145	2	H98144	hypothetical prote
14	93.5	12.0	645	2	F70825	probable pPE prote
15	93.5	12.0	1567	2	S11672	ice nucleation pro
16	92.5	11.9	151	2	S70787	curlin nucleator p
17	92.5	11.9	151	2	C90806	minor curlin subun
18	92.5	11.9	151	2	G85665	curlin minor chain
19	91.5	11.8	1034	2	JC2143	ice nucleation act
20	91.5	11.8	1258	2	JQ0188	ice nucleation pro
21	91	11.7	1751	2	G71518	hypothetical prote
22	90.5	11.7	151	2	JC6040	fimbrin protein ag
23	90.5	11.7	151	2	AH0635	nucleation compone
24	90.5	11.7	1053	2	B70987	probable pPE prote
25	89.5	11.5	1322	2	S07053	ice nucleation pro
26	89	11.5	575	2	S35327	protein kinase egg
27	87	11.2	251	2	D96010	hypothetical expor
28	87	11.2	434	2	E70768	hypothetical glyci
29	87	11.2	573	2	C86265	F3f19.21 protein -

```
30      87      11.2      590      1      A45621      leishmanolysin (EC
31      87      11.2      599      2      B42049      leishmanolysin (EC
32      87      11.2      599      2      A44951      leishmanolysin (EC
33      87      11.2      602      1      P40221      leishmanolysin (EC
34      87      11.2      646      1      S19916      leishmanolysin (EC
35      86.5     11.1     262      2      S00275      tail fiber protein
36      86.5     11.1     487      2      C70830      probable pPE prote
37      86.5     11.1     1028     2      A56038      DNA-binding protei
38      86.5     11.1     1213     2      F70675      obo protein - frui
39      86      11.1     582      2      F70675      probable pPE prote
40      85.5     11.0     940      2      D89723      protein F39D8.1b [
41      85.5     11.0     945      2      T21998      hypothetical prote
42      85.5     11.0     1210     2      A25547      ice nucleation pro
43      85      11.0     354      2      B70663      probable pPE prote
44      85      11.0     1052     2      A72959      conserved hypotet
45      85      11.0     1341     2      H98323      hypothetical prote
```

ALIGNMENTS

RESULT 1

JC6039
fimbrin protein agfa precursor - Salmonella enteritidis
C/Species: Salmonella enteritidis
C/Date: 31-Dec-1996 #sequence revision 31-Dec-1996 #text_change 08-Oct-1999
C/Accession: JC6039; PC6015; A44898
R/Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Baner, P.A.; Kay, W.W.
J. Bacteriol. 179, 662-667, 1996
A/Ritle: Salmonella enteritidis agfBAC operon encoding thin, aggregative fimbrae.
A/Reference number: JC6039; MUID:96146512; PMID:8550497
A/Accession: JC6039
A/Molecule type: DNA
A/Residues: 1-151 <COL>
A/Cross-references: GB:U43280; NID:gl1184712; PIDN:AAC43599.1; PID:gl1184714
A/Accession: PC6015
A/Molecule type: protein
A/Residues: 21-52 <CO2>
A/Experimental source: strain 27655-3b
A/Note: the authors translated the codon ACG for residue 44 as Ile
R/Collinson, S.K.; Emody, L.; Muller, K.H.; Trust, T.J.; Kay, W.W.
J. Bacteriol. 173, 4773-4781, 1991
A/Ritle: Purification and characterization of thin, aggregative fimbrae from Salmonella
A/Reference number: A44898; MUID:91310586; PMID:1677357
A/Contents: 27655
A/Accession: A44898
A/Status: preliminary
A/Molecule type: protein
A/Residues: 21-33 <CO3>
A/Note: sequence extracted from NCBI backbone (NCBIP:45936)
C/Genetics:
A/Gene: agfa
C/Function:
A/Description: major component of thin aggregative fimbrae
A/Note: fimbrae bind to fibronectin, plasminogen, tissue plasminogen activator
C/Keywords: fimbrin
F/1-20/Domain: signal sequence #status predicted <SIG>
F/21-151/Product: fimbrin protein agfa #status experimental <MAT>

Query Match 87.9%; Score 682; DB 2; Length 151;

Best Local Similarity 90.7%; Pred. No. 1.2e-50;

Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

Qy 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGSPDSTLSIYQYGSANAALAQ 60

1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGSPDSTLSIYQYGSANAALAQ 60

Qy 61 SDARKSETTITQSGYNGADVGOGADNYQLVTRVVTTHMAHADOWNAKNSDITTVQYGG 120

61 SDARKSETTITQSGYNGADVGOGADNYQLVTRVVTTHMAHADOWNAKNSDITTVQYGG 120

Qy 121 NNAALVNTQADSVMVRQVGFNNATANQY 151

Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 2

AI0635

major curlin chain precursor [imported] - Salmonella enterica subsp. enterica serovar Typhimurium

C:Species: Salmonella enterica subsp. enterica serovar Typhi

A:Note: this species has also been called Salmonella typhi

C>Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002

C:Accession: AI0635

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moulle, S.; O'Garra, P.

Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.

A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhimurium

A:Reference number: AB0502; MUID:21534947; PMID:11677608

A:Accession: AI0635

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-151 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD08268.1; PID:gi6502315; GSPDB:GN00176

C:Genetics:

A:Gene: STY1181

Query Match 87.9%; Score 682; DB 2; Length 151;

Best Local Similarity 90.7%; Pred. No. 1.2e-50;

Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQGGGNGHNGGNSGDPSTLSIYQGSANAALALQ 60

DB 1 MKLLKVAAPAAIVVSGSALAGVVPQGGGNGHNGGNSGDPSTLSIYQGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGADVGGADNYDQVLTVRVVTHEMAHADOWNAKNSDITVQYGG 120

DB 61 SDARKSETTITQSGYNGADVGGADNYDQVLTVRVVTHEMAHADOWNAKNSDITVQYGG 120

QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 3

S70788

curlin protein csgA precursor - Escherichia coli (strain K-12)

N:Alternate names: csgA protein; major curlin protein

C:Species: Escherichia coli

C>Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 01-Mar-2002

C:Accession: S70788; G64846; S31202; S34560; S34559

R:Hammar, M.; Arngvist, A.; Bian, Z.; Olsen, A.; Normark, S.

Mol. Microbiol. 18, 661-670, 1995

A:Title: Expression of two csg operons is required for production of fibronectin- and C

A:Reference number: S70788; MUID:96414458; PMID:8817489

A:Accession: S70788

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-151 <HAM>

A:Cross-references: EMBL:X90754; NID:gi1147558; PIDN:CAA62282.1; PID:gi1147564

A:Experimental source: strain K12, substrain W3110

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co

A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: G64846

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-151 <BLAT>

A:Cross-references: GB:AE000205; GB:U00096; NID:gi1787265; PIDN:AAC74126.1; PID:gi1787279;

A:Experimental source: strain K-12, substrain MG1655

R:Olsen, A.; Arngvist, A.; Hammar, M.; Sukupolvi, S.; Normark, S.

Mol. Microbiol. 7, 523-536, 1993

A:Title: The Eps sigma factor relieves H-NS-mediated transcriptional repression of csgA

A:Reference number: S31202; MUID:93211294; PMID:8459772

A:Accession: S31202

A:Molecule type: DNA

A:Residues: 1-6,'V',8-151 <OLS1>

A:Cross-references: EMBL:L04979

A:Accession: S34560

A:Molecule type: protein

A:Residues: 21-42;44-50 <OLS2>

R:Olsen, A.N.; Arngvist, A.M.

submitted to the EMBL Data Library, October 1992

A:Reference number: S34559

A:Accession: S34559

A:Molecule type: DNA

A:Residues: 1-133,'RQRDSGLW' <OLS3>

A:Cross-references: EMBL:L04979; NID:g290424; PIDN:AAA23616.1; PID:g290425

A:Experimental source: strain K-12, substrain W3110

C:Genetics:

A:Gene: csgA

A:Map position: 23.15

C:Function:

A:Description: major component of wild-type curli; interaction between CsgA and CsgB tri

A:Note: curli are thin, coiled fibers expressed on the surface of Escherichia coli that

and H-kininogen; in the absence of CsgA, CsgB can self-assemble into polymers

F:1-20/Domain: signal sequence #status predicted <SIG>

F:21-151/Product: curlin #status experimental <MAT>

Query Match 67.1%; Score 521; DB 2; Length 151;

Best Local Similarity 69.5%; Pred. No. 4.6e-37;

Matches 105; Conservative 17; Mismatches 29; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQGGGNGHNGGNSGDPSTLSIYQGSANAALALQ 60

DB 1 MKLLKVAAPAAIVVSGSALAGVVPQGGGNGHNGGNSGDPSTLSIYQGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGADVGGADNYDQVLTVRVVTHEMAHADOWNAKNSDITVQYGG 120

DB 61 TDARNSDLTITQGGGNGADVGGSDSSIDITQGFNGSATLDOWNGKNSDITVQYGG 120

QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

DB 121 GNGAAVDQTASNSVNVTVQVGFNNATAHQY 151

RESULT 4

D90806

curlin major subunit CsgA [imported] - Escherichia coli (strain O157:H7, substrain RIMD

C:Species: Escherichia coli

C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001

C:Accession: D90806

R:Hayaishi, T.; Makino, K.; Onishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.

Gasawara, N.; Yasunaga, T.; Kihara, S.; Shiba, T.; Hattori, M.; Shingawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene

A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: D90806

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-152 <HAY>

A:Cross-references: GB:BA000007; PIDN:BA034843.1; PID:gi13360880; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RIMD 050952

C:Genetics:

A:Gene: ECs1420

Query Match 64.4%; Score 499.5; DB 2; Length 152;

Best Local Similarity 67.8%; Pred. No. 3e-35;

Matches 103; Conservative 18; Mismatches 30; Indels 1; Gaps 1;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQGGGNGHNGGNSGDPSTLSIYQGSANAALAL 59

DB 1 MKLLKVAAPAAIVVSGSALAGVVPQGGGNGHNGGNSGDPSTLSIYQGSANAALAL 60

QY 60 QSDARKSETTITQSGYNGADVGGADNYDQVLTVRVVTHEMAHADOWNAKNSDITVQYGG 119

```

Db 61 QADARNSLTTTQGGNGADVGCGSDSSIDLTORFGNSATLDQWNGKDSHTVKQFG 120
QY 120 GNNAAALVNOTASDSSVMVROVGFQGNNTATNQY 151
Db 121 CGNGAAVDQTASNTVNTVQVGFQGNNTATNQY 152

RESULT 5
H85665
hypothetical protein csgA [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: H85665
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: H85665
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-152 <SOT>
A:Cross-references: GB:AB001574; NID:G12514574; PIDN:AAG55788.1; GSPDB:GN00145; UWGP:Z16
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: csgA

Query Match 64.4%; Score 499.5; DB 2; Length 152;
Best Local Similarity 67.8%; Pred. No. 3e-35;
Matches 103; Conservative 18; Mismatches 30; Indels 1; Gaps 1;

QY 1 MKLLKVAFAIVVSGSALAGVPOW-GGGNHNGGNSGPDSTLSIYQYGSNAALAL 59
Db 1 MKLLKVAIAAIV7SGSALAGVPPQYGGGGHGGGNGSGNSGPNSELNIYQYGGNSALAL 60

60 QSDARKSETTTTQSGYNGADVGADGADNYDQLVTRVVTTHMAHADOWNAKNSDITVQY 119
QY 119 QSDARKSETTTTQSGYNGADVGADGADNYDQLVTRVVTTHMAHADOWNAKNSDITVQY 119
Db 61 QADARNSLTTTQGGNGADVGCGSDSSIDLTORFGNSATLDQWNGKDSHTVKQFG 120

QY 120 GNNAAALVNOTASDSSVMVROVGFQGNNTATNQY 151
Db 121 CGNGAAVDQTASNTVNTVQVGFQGNNTATNQY 152

RESULT 6
cnjB protein - Tetrahymena thermophila
C:Species: Tetrahymena thermophila
C:Date: 19-Mar-1997 #sequence_revision 18-Jul-1997 #text_change 07-Dec-1999
C:Accession: S42136; S42135; S03650
R:Taylor, F.M.; Martindale, D.W.
submitted to the EMBL Data Library, October 1992
A:Reference number: S42136
A:Accession: S42136
A:Molecule type: DNA
A:Residues: 1-1748 <TAY>
A:Cross-references: EMBL:L03710; NID:G161751; PID:G161752
R:Taylor, F.M.; Martindale, D.W.
Nucleic Acids Res. 21, 4610-4614, 1993
A:Title: Retroviral-type zinc fingers and glycine-rich repeats in a protein encoded by c
A:Reference number: S42135; MUID:94051569; PMID:8233798
A:Accession: S42135
A:Molecule type: DNA
A:Residues: 1154-1174; 1179-1198; 1233-1252; 1285-1293; 1316-1326; 1331-1341; 1343-1
A:Cross-references: EMBL:L03710
R:Martindale, D.W.; Taylor, F.M.
Nucleic Acids Res. 16, 2189-2201, 1988
A:Title: Multiple introns in a conjugation-specific gene from Tetrahymena thermophila.
A:Reference number: S03650; MUID:88189811; PMID:3357771
A:Accession: S03650
A:Molecule type: DNA
A:Residues: 236-250, '1', 252-255, 'N', 257-773 <MAR>
```

```

A:Cross-references: EMBL:X06462
C:Genetics:
A:Gene: cnjB
A:Genetic code: SGCS
A:Introns: 85/3; 136/1; 157/3; 201/2; 290/2; 327/3; 499/1; 573/2; 607/3; 708/3; 777/3;
C:Keywords: zinc finger
F:1164-1450/Region: glycine-rich
F:1451-1464/Region: zinc finger CCHC motif
F:1478-1491/Region: zinc finger CCHC motif
F:1501-1514/Region: zinc finger CCHC motif
F:1530-1543/Region: zinc finger CCHC motif
F:1555-1568/Region: zinc finger CCHC motif
F:1579-1592/Region: zinc finger CCHC motif
F:1602-1615/Region: zinc finger CCHC motif
F:1626-1748/Region: glycine-rich

Query Match 14.0%; Score 109; DB 2; Length 1748;
Best Local Similarity 33.6%; Pred. No. 0.39;
Matches 37; Conservative 11; Mismatches 32; Indels 30; Gaps 6;

QY 25 QWGGGNGHNGG---GNSSGPDSTLSIYQYGSNAALALQSDARKSETTIT---QSGYCN 77
Db 1640 QFGGGNSNGSGSWGTSSTGSDWN-----CQSNVQESTTTSSGGWGGSGGSGN 1685

QY 78 GADVGCGA-DNYDQLVTRVVTTHMAHADOWNAKNSDITVGO--YGGNAA 124
Db 1686 QTGGGWSNDNQO-----QNETGGGCGGSSNQTNNESSWGSNNQA 1729

RESULT 7
E95965
hypothetical glycine-rich protein [imported] - Sinorhizobium meliloti (strain 1021) mag
C:Species: Sinorhizobium meliloti
C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C:Accession: E95965
R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Herna
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A:Title: The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing endo
A:Reference number: A95842; MUID:21396508; PMID:11481431
A:Accession: E95965
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2174 <KUR>
A:Cross-references: GB:AL591985; PIDN:CAC49389.1; PID:G15140875; GSPDB:GN00167
A:Experimental source: strain 1021, megaplasmid pSymB
R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler
Pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.
L.; Hyman, R.W.; Jones, I.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure
hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A:Reference number: A96039; MUID:21368234; PMID:11474104
A:Contents: annotation
C:Genetics:
A:Gene: SMB21548
A:Genome: plasmid

Query Match 14.0%; Score 109; DB 2; Length 2174;
Best Local Similarity 27.0%; Pred. No. 0.49;
Matches 40; Conservative 20; Mismatches 52; Indels 36; Gaps 7;

QY 11 AIVVGSALAGVVPQ--WGCGNHNGGNSGPDSTLSIYQYGS-----ANAA----- 56
Db 693 AATAGAGAVGILAQISGGGNG--GGNATGADAGFGSGFQGGGGGGGYANTANVGPK 749

QY 57 -LALQSDARKSETTTTQSGYNGADVGCGADNYDQLVTRVVTTHMAHADOWNAKNSDITV 115
Db 750 GUTLTTQGHAGIVAQVS-VGGGGGTGTTASSISAGI-----GFTASVAV 793

QY 116 GQYGGNAA--LVNOTASDSSVMVROVQ 141
Db 794 GGTGGNGGAGGVSVELTDSAIRTGQGG 821
```



```

Db      383 GSTGTAGADSTL-IAGYGSTQTA-----GGESLTL-AGYGTQTARQGSDDITAGYGS 432
QY      89 -----DQLVTRVVTHEMAHAD-----QWNAKNSDITVGGYG-----GNNAAALV 126
Db      433 TGTAGADSTLIAGYGSTQTSQSDSSLTAGYGSTQTARKGSDITAG-YGSTGTAGSDSLI 491
QY      127 -----NOTASDSSVMVRQVFGNNATANQ 150
Db      492 AGYGTQTAGSESSLT--AGYGTQTAAQ 518

```

Search completed: August 2, 2004, 14:56:24
 Job time : 10.4 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:36:12 ; Search time 5.3 seconds
(without alignments)

1483.508 Million cell updates/sec

Title: US-09-543-407-22

Perfect score: 776

Sequence: 1 MKLKVAFAAIVVSGSALA.....DSSVMVRQVGFNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	682	87.9	151	1	CSGA_SALTY
2	521	67.1	151	1	CSGA_ECOLI
3	499.5	64.4	152	1	CSGA_ECO57
4	93.5	12.0	1567	1	ICEN_XANGT
5	92.5	11.9	151	1	CSGB_ECOLI
6	91.5	11.8	1034	1	ICEN_PANAN
7	91.5	11.8	1258	1	ICEN_ERWHE
8	91	11.7	1754	1	PMPB_CHLTR
9	90.5	11.7	151	1	CSGB_SALTY
10	90.5	11.7	151	1	ICBA_PANAN
11	89.5	11.5	1322	1	ICBA_PANAN
12	87	11.2	491	1	YK98_MYCTU
13	87	11.2	590	1	GP63_LEIDO
14	87	11.2	599	1	GP63_LEICH
15	87	11.2	602	1	GP63_LEIVA
16	87	11.2	646	1	GP63_LEIME
17	86.5	11.1	262	1	V338_BPT2
18	86.5	11.1	487	1	Y442_MYCTU
19	86.5	11.1	548	1	CEAK_ECOLI
20	86.5	11.0	1028	1	OVO_DROME
21	85.5	11.0	1210	1	ICEN_PSEFL
22	85	11.0	1778	1	N189_SCHPO
23	84.5	10.9	495	1	P033_MOUSE
24	84.5	10.9	760	1	YBTL_ECOLI
25	84.5	10.9	1211	1	YBTL_ECOLI
26	83.5	10.8	497	1	P033_RAT
27	83	10.7	1140	1	YW96_YEAST
28	82	10.6	500	1	P033_HUMAN
29	82	10.6	678	1	YF48_MYCTU
30	81.5	10.5	392	1	HWE1_HUMAN
31	81.5	10.5	737	1	YD70_MYCFN
32	81.5	10.5	2038	1	FSH_DROME
33	81	10.4	165	1	GRF1_ORISA

34	80.5	10.4	874	1	ALB6_AZOVI
35	80.5	10.4	1196	1	ICEV_PSEX
36	80	10.3	204	1	CORA_MEDSA
37	80	10.3	362	1	ALB1_STACP
38	80	10.3	429	1	DR48_YEAST
39	80	10.3	443	1	Y978_MYCTU
40	80	10.3	543	1	YP91_MYCTU
41	79.5	10.2	396	1	PER_DROPV
42	78.5	10.1	346	1	RO21_XENLA
43	78.5	10.1	1148	1	ICEK_PSEX
44	78.5	10.1	1317	1	N145_YEAST
45	78	10.1	362	1	P35_MYCPE

ALIGNMENTS

RESULT 1				
CSGA_SALTY	STANDARD;	PRT;	151 AA.	
ID	CSGA_SALTY	STANDARD;	PRT;	151 AA.
AC	P5225;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Major curlin subunit precursor (Fimbrin SEF17).			
GN	CSGA OR AGFA OR STM1144 OR STY1181 OR T1776.			
OS	Salmonella typhimurium,			
OS	Salmonella typhi, and			
OS	Salmonella enteritidis.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OC	Enterobacteriaceae; Salmonella.			
OX	NCBI_TaxID=602, 601, 592;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=S.typhimurium; STRAIN=SP-11;			
RX	MEDLINE=98117058; PubMed=9457880;			
RA	Romling U., Bian Z., Hammar M., Sierralta W.D., Normark S.;			
RT	"Curli" fibers are highly conserved between Salmonella typhimurium and			
RT	Escherichia coli with respect to operon structure and regulation.;"			
RL	J. Bacteriol. 180:722-731(1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=S.typhimurium; STRAIN=LT2 / SGSCI412 / ATCC 700720;			
RX	MEDLINE=21534948; PubMed=11677609;			
RA	McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,			
RA	Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,			
RA	Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,			
RA	Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,			
RA	Waterston R., Wilson R.K.;			
RT	"Complete genome sequence of Salmonella enterica serovar Typhimurium			
RT	LT2.;"			
RL	Nature 413:852-856(2001).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=S.typhi; STRAIN=CT18;			
RX	MEDLINE=21534947; PubMed=11677608;			
RA	Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,			
RA	Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,			
RA	Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,			
RA	Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,			
RA	Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,			
RA	Krog A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,			
RA	Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,			
RA	Whitehead S., Bartell B.G.;			
RT	"Complete genome sequence of a multiple drug resistant Salmonella			
RT	enterica serovar Typhi CT18.;"			
RL	Nature 413:848-852(2001).			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=S.typhi; STRAIN=Ty2 / ATCC 700931;			
RX	MEDLINE=22531367; PubMed=12644504;			
RA	Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,			
RA	Burland V., Kodoyanni V., Schwartz D.C., Blattner P.R.;			

Q92fh0	azotobacter
Q33479	pseudomonas
Q07202	medicago sa
Q05156	staphylococ
P18999	saccharomyc
Q10340	mycobacteri
Q50630	mycobacteri
P91698	drosophila
P51989	xenopus lae
Q30611	pseudomonas
P49687	saccharomyc
Q50367	mycoplasma

```

RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RL J. Bacteriol. 185:2330-2337(2003).
RN [5]
RP SEQUENCE FROM N.A.
RC SPECIES=S. enteritidis; STRAIN=27655-3B;
RX MEDLINE=96146512; PubMed=8550497;
RA Collinson S.K., Clouthier S.C., Doran J.L., Baner P.A., Kay W.W.;
RT "Salmonella enteritidis agfBAC operon encoding thin, aggregative
fimbriae.";
RN J. Bacteriol. 178:662-667(1996).
RN [6]
RP SEQUENCE OF 21-151 FROM N.A.
RC SPECIES=S. enteritidis; STRAIN=27655-3B;
RX MEDLINE=94013373; PubMed=8104955;
RA Doran J.L., Collinson S.K., Burian J., Santos G., Todd E.C.D.,
Munro C.K., Kay C.M., Baner P.A., Peterkin P.I., Kay W.W.;
RT "DNA-based diagnostic tests for Salmonella species targeting agfA,
the structural gene for thin, aggregative fimbriae.";
RN J. Clin. Microbiol. 31:2263-2273(1993).
RN [7]
RP SEQUENCE OF 21-33.
RC SPECIES=S. enteritidis; STRAIN=27655-3B;
RX MEDLINE=91310586; PubMed=1677357;
RA Collinson S.K., Emedy L., Mueller K.-M., Trust T.J., Kay W.W.;
RT "Purification and characterization of thin, aggregative fimbriae from
Salmonella enteritidis.";
RN J. Bacteriol. 173:4773-4781(1991).
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN.
CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ002301; CA005317.1; -
DR EMBL; AF008749; RAL20074.1; -
DR EMBL; AL627269; CAD08268.1; -
DR EMBL; AF016840; AAC09399.1; -
DR EMBL; U43280; AAC43599.1; -
DR PIR; JC6039; JC6039.
DR StyGene; SGI0608; CSGA.
KW Fimbria; Signal; Complete proteome.
FT SIGNAL 1 20
FT CHAIN 21 151 MAJOR CURLIN SUBUNIT.
FT CONFLICT 134 151 SVMVRQVGFNNATANQY --> DSVTQVAS (IN
FT REF. 6).
SQ SEQUENCE 151 AA; 15305 MW; B7DAC0D16B621359 CRC64;
Query Match 87.9%; Score 682; DB 1; Length 151;
Local Similarity 90.7%; Pred. No. 2.2e-50;
Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;
QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNNHGGNSGGPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNNHGGNSGGPDSTLSIYQYGSANAALALQ 60
QY 61 SPARKSETTITSGYNGGADVGCGADNDYDQLVTRVVTHEMADOWNAKNSDITVGYGG 120
DB 61 SPARKSETTITSGYNGGADVGCGADNDYDQLVTRVVTHEMADOWNAKNSDITVGYGG 120
QY 121 NNAALVNQATSDSSVMVRQVGFNNATANQY 151
DB 121 NNAALVNQATSDSSVMVRQVGFNNATANQY 151

```

```

RESULT 2
CSGA_ECOLI
ID _CSGA_ECOLI STANDARD; PRT; 151 AA.
AC P28307;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Major curlin subunit precursor.
GN CSGA OR B1042.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / W3110;
RX MEDLINE=93211294; PubMed=8459772;
RA Olsen A., Arngvist A.;
RT "The Kps sigma factor relieves H-NS-mediated transcriptional
repression of csGA, the subunit gene of fibronectin-binding curli in
Escherichia coli.";
RN Mol. Microbiol. 7:523-536(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MC4100;
RX MEDLINE=96414488; PubMed=8817489;
RA Hammar M., Arngvist A., Bian Z., Olsen A., Normark S.;
RT "Expression of two csG operons is required for production of
fibronectin- and congo red-binding curli polymers in Escherichia coli
K-12.";
RN Mol. Microbiol. 18:661-670(1995).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MGL655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RN Science 277:1453-1474(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 12.7-28.0 min region on the linkage map.";
RN DNA Res. 3:137-155(1996).
RN [5]
RP SEQUENCE OF 21-40.
RC STRAIN=K12 / YNML;
RX MEDLINE=93023873; PubMed=1357528;
RA Arngvist A., Olsen A., Pfeifer J., Russell D.G., Normark S.;
RT "The Crl protein activates cryptic genes for curli formation and
fibronectin binding in Escherichia coli HB101.";
RN Mol. Microbiol. 6:2443-2452(1992).
RN [6]
RP SEQUENCE OF 21-31.
RC MEDLINE=91310586; PubMed=1677357;
RA Collinson S.K., Emedy L., Trust T.J., Kay W.W.;
RT "Purification and characterization of thin, aggregative fimbriae from
Salmonella enteritidis.";
RN J. Bacteriol. 173:4773-4781(1991).
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN.

```

```

[3]
RN RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasaki C., Ogasawara N., Yasunaga T.,
RA Kihara S., Shiba T., Hattori M., Shingawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12."
RL DNA Res. 8:11-22(2001).
CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN.
CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR ENBL; A8275733; AAK53212.1; -
DR ENBL; A8005315; AAG55788.1; -
DR ENBL; APO02554; BAB34843.1; -
DR PIR; D90806; D90806.
DR PIR; H85665; H85665.
DR KW Fibria; Signal; Complete proteome.
FT SIGNAL 1 20 BY SIMILARITY.
FT CHAIN 21 152 MAJOR CURLIN SUBUNIT.
FT FT
SQ SEQUENCE 152 AA; 15099 MW; EE2D2D94DDE91243 CRC64;

Query Match 64.4%; Score 499.5; DB 1; Length 152;
Best Local Similarity 67.8%; Pred. No. 3.8e-35;
Matches 103; Conservative 18; Mismatches 30; Indels 1; Gaps 1;

QY 1 MKLLKVAFAAIIVVSGSALAGVVPW-GGGGNHGGNGSGDPSTLSIYQYGSANAAL 59
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MKLLKVAFAAIIVVSGSALAGVVPVQVGGGNGHGGNGSGPNSLNIVYQGGNSALAL 60
QY 60 QSDARKSETITQSGYNGADYVGGQADNYDQLVTRVVTTHMAHADQWNAKNSDITVGYG 119
Db 61 QADARNSDLTITQHGCGGADYVGGQSDSSIDLITQRFGNSATLDQWNGKDSHTVTKQFG 120
QY 120 GNNAAVLNQTASDSSVNVTVQVFGNNATAHQY 151
Db 121 GNGRAVDQTASNSTVNVTVQVFGNNATAHQY 152

RESULT 4
ICEN_XANCT STANDARD; PRT; 1567 AA.
ID ICEN_XANCT
AC F18127;
DT 01-NOV-1990 (Rel. 16, Created)
DD 01-NOV-1990 (Rel. 16, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ice nucleation protein.
DE GN
DE INAX.
OS Xanthomonas campestris (pv. translucens).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OC NCBI_TaxID=343;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=X56S;
RX MDLINE=91080859; PubMed=2259339;
RA Zhao J., Orser C.S.;
RT "Conserved repetition in the ice nucleation gene inax from
RT Xanthomonas campestris pv. translucens."
RL Mol. Gen. Genet. 223:163-166(1990).
```



```

QY 58 ALQSDARKSETTITGSGYGN-----GADVCGG-----ADNYDQLVTRVVTHEMAHAD- 104
Db 219 -----GESSQWAGYSGTQGMKGDLTAGYSGTGTAGDSSSLIAGYSGTGTAGDS 270
QY 105 -----QWNAKNSDITVQYGGNNAALVNQATASDSSVMVRQVGFNNATANQ 150
Db 271 SLTAGYGTQTAQKGSDLTAG-YGSTGTA-----GADSSLI-----AGYSGTGTAGE 316

RESULT 8
PMPB_CHLTR
ID PMPB_CHLTR STANDARD; PRT; 1754 AA.
AC O84418;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable outer membrane protein pmpB precursor (Polymorphic membrane
protein B).
GN PMPB OR C7413.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D/UW-3/Cx;
RX MEDLINE=99000809; PubMed=9784136;
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis."
RL Science 282:754-759(1998).
CC -! SUBCELLULAR LOCATION: Cell wall surface (elementary bodies)
CC (potential).
CC -! SIMILARITY: Belongs to the PMP outer membrane protein family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB001314; AAC68010.1; ALT_INIT.
CC PHC1-2DPAGE; O84418; -.
CC InterPro; IPR003368; Chlamydia_PMP.
CC Pfam; PF02415; Chlamydia_PMP; 4.
CC TIGRfam; TIGR01376; POMP-repeat; 18.
KW Outer membrane; Signal; Multigene family; Complete proteome.
FT SIGNAL 1 14 POTENTIAL.
FT CHAIN 15 1754 PROBABLE OUTER MEMBRANE PROTEIN PMPB.
SQ SEQUENCE 1754 AA; 183317 MW; 969CF8D8D36185D CRC64;

Query Match 11.7%; Score 91; DB 1; Length 1754;
Best Local Similarity 33.1%; Pred. No. 7.3;
Matches 40; Conservative 10; Mismatches 59; Indels 12; Gaps 5;

QY 15 GSALAGVVPQWGGGN--HNGGNSGGPDSTLSIYQYGSANAALALQSDARKSETTIT- 71
Db 397 NADAWASSPQSGGATTVNSGDSGSDSDTSETVPATKGG-GLYTDKNLSITNITG 455
QY 72 --QSGYNGADVQGANVDQLVTRVTHEMAHADQWNAKNSDITVQY- YGNNNAALVNQ 128
Db 456 IIEIANNKATDVGGA-----YVXGTLTENSRLQFLKNSGDKOGGIYGEDNITLSNL 510
QY 129 T 129
Db 511 T 511

RESULT 9
CSGB_SALTY
ID CSGB_SALTY STANDARD; PRT; 151 AA.

```

```

ID CSGB_SALTY STANDARD; PRT; 151 AA.
AC Q827M3;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Minor curlin subunit precursor.
GN CSGB OR STYL180 OR T1777.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Hague A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrrell B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18."
RL Nature 413:848-852(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodyanmi V., Schwartz D.C., Blattner F.R.;
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RT and CT18."
RL J. Bacteriol. 185:2330-2337(2003).
CC -! FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
CC CURLIN MONOMERS.
CC -! SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AL627269; CAD08267.1; -.
CC DR EMBL; AL6016940; AAC69400.1; -.
CC KW Fimbria; Signal; Complete proteome.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 151 MINOR CURLIN SUBUNIT.
SQ SEQUENCE 151 AA; 16254 MW; 161C543265573495 CRC64;

Query Match 11.7%; Score 90.5; DB 1; Length 151;
Best Local Similarity 28.8%; Pred. No. 0.53;
Matches 34; Conservative 19; Mismatches 46; Indels 19; Gaps 6;

QY 51 GSANAALALQSDARKSE-----TTTQSGYNGADVQY-GADNYDQL-----VTRVWTH 98
Db 14 GAGCIATATNYDLARSEYNAFVNELSKSFFNQAAIGQVGTDSARVROEGSKLLSVISQ 73
QY 99 E-----MAHADQWNAKNSDIT-VQYCGNNAALVNQATASDSSVMVRQVGFNNATANQ 151
Db 74 EGENNRKVDQ--AGNYNFAYIEQTGNANDASISQAYGNSAAITQKSGNKNITQY 129

RESULT 10
CSGB_SALTY
ID CSGB_SALTY STANDARD; PRT; 151 AA.

```

```

AC P55226;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Minor curlin subunit precursor (Pimbrin SEF17 minor subunit).
GN CSGS OR AGFB OR STM1143.
OS Salmonella typhimurium, and
OS Salmonella enteritidis
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602, 592;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=SR-11;
RX MEDLINE=98117058; PubMed=9457880;
RA Rowling U., Bian Z., Hammar M., Sierralta W.D., Normark S.;
RT "Curli fibers are highly conserved between Salmonella typhimurium and
RT Escherichia coli with respect to operon structure and regulation.";
RL J. Bacteriol. 180:722-731(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=S.enteritidis; STRAIN=27655-3B;
RX MEDLINE=96146512; PubMed=8550497;
RA Collinson S.K., Clouthier S.C., Doran J.L., Bansen P.A., Kay W.W.;
RT "Salmonella enteritidis agfBAC operon encoding thin, aggregative
RT fimbriae.";
RL J. Bacteriol. 178:662-667(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=27655-3B;
RX MEDLINE=96146512; PubMed=8550497;
RA Collinson S.K., Clouthier S.C., Doran J.L., Bansen P.A., Kay W.W.;
RT "Salmonella enteritidis agfBAC operon encoding thin, aggregative
RT fimbriae.";
RL J. Bacteriol. 178:662-667(1996).
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
CC CURLIN MONOMERS.
CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AJ002301; CAA05316.1; -
CC EMBL; AE008745; AAL20073.1; -
CC EMBL; U43280; AAC43598.1; -
CC FIR; JC6040; JC6040.
CC Dr StyGene; SG10509; csGB.
CC Fimbria; Signal; Complete proteome.
FT SIGNAL 1 21
FT CHAIN 22 151 MINOR CURLIN SUBUNIT.
FT SEQUENCE 151 AA; 16182 MW; CQPC543086DD361D CRC64;
CC
Query Match 11.7%; Score 90.5; DB 1; Length 151;
Best Local Similarity 27.5%; Pred. No. 0.53;
Matches 36; Conservative 17; Mismatches 39; Indels 39; Gaps 5;
CC
QY 10 AAIIVVGSALAGVYQWGGGNNHGGNSGPDSTLSIYQYGSANAALQSDARKSETT 69
DB 58 ARVREGSKLLSVISQ--EGGNRAKVQDQGNYPAYIEQIGNAN-----DAS 103
QY 70 ITQSGYGNADVGQ--GADNYDQLVTRVVTHEMAHADQNNAKNSDITVQYQGNNAALVQ 128

```

```

Db 104 ISQSAVGNAAIIQKSGN-----KANIT--QYGTQKTAVVQ 139
QY 129 TASDSSVMVRQ 139
Db 140 KOSHVAIRVTQ 150
RESULT 11
ID ICEA_PANAN STANDARD; PRT; 1322 AA.
AC P20469;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ice nucleation protein inaA.
GN INAA.
OS Pantoea ananas (Erwinia uredovora).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pantoea.
OX NCBI_TaxID=553;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90092494; PubMed=2599095;
RA Abe K.; Watabe S.; Emori Y.; Watanabe M.; Arai S.;
RT "An ice nucleation active gene of Erwinia ananas. Sequence similarity
RT to those of Pseudomonas species and regions required for ice
RT nucleation activity.";
RL FEBS Lett. 258:297-300(1989).
CC -!- FUNCTION: Ice nucleation proteins enable bacteria to nucleate
CC crystallization in supercooled water.
CC -!- SUBCELLULAR LOCATION: Outer membrane (By similarity).
CC -!- DOMAIN: CONTAINS MANY IMPERFECT REPEATS OF THE CONSENSUS
CC OCTAPEPTIDE A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A
CC REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.
CC -!- SIMILARITY: Belongs to the bacterial ice nucleation protein
CC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X17316; CAA35194.1; -
CC FIR; S07053; S07053.
CC HSSP; P06620; IINA.
CC InterPro; IPR000258; Ice_nucleatn.
CC Pfam; PF00818; Ice_nucleation; 69.
CC PRINTS; PR00327; ICENUCLEATN.
CC PROSITE; PS00314; ICE_NUCLEATION; 49.
CC Ice_nucleation; Repeat; Outer membrane.
FT DOMAIN 162 1281 OCTAPEPTIDE PERIODICITY.
FT SEQUENCE 1322 AA; 131094 MW; 89B0EE24AA837039 CRC64;
CC
Query Match 11.5%; Score 89.5; DB 1; Length 1322;
Best Local Similarity 26.8%; Pred. No. 7.1;
Matches 34; Conservative 18; Mismatches 26; Indels 49; Gaps 7;
CC
QY 34 GGNSSGPDSTLSIYQYGSANAALQSDARKSETTITQSGYGNADVGQGDADNYDQLVT 93
DB 933 GSTSTAGPSSL-IAGYGSTQTA-----GYSILTAGY----- 965
QY 94 RVVTHEMAHADQNNAKNSDITVQYQV-----GNNALV-----NOTASDSSVMVRQVGG 143
DB 966 -----STQTQGNESDLTTG-YGSTSTAGYESSLIAGYGSTOTASPFTLM--AGYG 1013
QY 144 NNATANQ 150
DB 1014 SSQTARE 1020

```

```

RESULT 12
YK98 MYCTU
ID YK98 MYCTU STANDARD; PRT; 491 AA.
AC Q10707;
DT 01-OCT-1996 (Rel. 34, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical glycoprotein Rv2098c/MT2159/MB2125c.
GN Rv2098C OR MT2159 OR MT2159 OR MT2159 OR MB2125C.
OS Mycobacterium tuberculosis, and
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773, 1765;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=H37Rv;
RX MEDLINE=98293987; PubMed=934230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=1218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL J. Bacteriol. 184:5479-5490(2002).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=M.bovis; STRAIN=AF122/97;
RX MEDLINE=22709107; PubMed=12788972;
RA Garnier T., Eigmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Fryer M., Duttoy S., Gronin S., Lacroix C., Monsepe C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
CC -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS
CC SUBFAMILY
CC -!- CAUTION: Ref.1 sequence differs from that shown due to a
CC frameshift in position 59. Ref.1 sequence has been checked by
CC authors in Ref.1 and they report that no errors have been found.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; 273956; CAAG8228.1; ALT FRAME.
CC EMBL; AE007065; RAA46440.1; --
CC EMBL; BX248341; CAD96978.1; --
CC TIGR; MT2159; --
CC TubercuList; Rv2098c; --
CC InterPro; IPR000084; PE_region.
CC Pfam; PF00934; PE; 1.

```

```

DR ProDom; PD001223; PE_region; 1.
KW Hypothetical protein; Complete proteome.
FT CONFLICT 312 G -> GG (IN REF. 1).
SQ SEQUENCE 491 AA; 41979 MW; 12C8630C59CA0C13 CRC64;

Query Match 11.2%; Score 87; DB 1; Length 491;
Best Local Similarity 28.2%; Pred. NO. 3.8;
Matches 31; Conservative 8; Mismatches 45; Indels 26; Gaps 4;

QY 20 AGVVPMQGGGCGG-----HNGGGNSGPDSTLSIVYQGSANAALQSDARKSETTTQSG 74
DB 222 AGLIGHGAGGCGGCGGCGGSGKAGSGGSGFGGAGGGLL----- 264
QY 75 YGNADVGGQADNYDQIVRVVTHEMAHADQWNAKNSD---ITVGYGNG 121
DB 265 YNGGAAGSGGGGAGD-AGTGVSDDGFAGLGGSGGGGAGLIGVGGGCGN 313

RESULT 13
GP63 LEIDO STANDARD; PRT; 590 AA.
ID GP63 LEIDO
AC P23223;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Leishmanolysin precursor (EC 3.4.24.36) (Cell surface protease)
DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface
DE endopeptidase).
DN GP63.
OS Leishmania donovani.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5661;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LV95;
RX MEDLINE=92107220; PubMed=1762629;
RA Webb J.R., Button L.L., McMaster R.W.;
RT "Heterogeneity of the genes encoding the major surface glycoprotein
RT of Leishmania donovani.";
RL Mol. Biochem. Parasitol. 48:173-184(1991).
CC -!- FUNCTION: Has an integral role during the infection of macrophages
CC in the mammalian host.
CC -!- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and
CC P1' and basic residues at P2 and P3'. A model nonapeptide is
CC cleaved at -Ala-Tyr-I-Leu-Lys-Lys-.
CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -!- SIMILARITY: Belongs to peptidase family M8.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M60048; AAA23244.1; --
CC HSPB; P08148; 1LML.
CC MEROPS; M08.001; --
CC InterPro; IPR006025; Pept_M_Zn_BS.
CC InterPro; IPR001577; Peptidase_M8.
CC Pfam; PF01457; Peptidase_M8; 1.
CC PRINTS; PR00782; LSHMANOLYSIN.
CC PROSITE; PS00142; ZINC_PROTEASE; 1.
CC Hydrolase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;
CC Zymogen; Signal; Cell adhesion; GPI-anchor; Lipoprotein.
FT SIGNAL 1 39 POTENTIAL.
FT PROPEP 40 87 ACTIVATION PEPTIDE.
FT CHAIN 88 565 LEISHMANOLYSIN.
FT PROPEP 566 590 REMOVED IN NATURE FORM (BY SIMILARITY).
FT METAL 251 251 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 252 252 BY SIMILARITY.

```



```

FT METAL 255 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 321 ZINC (CATALYTIC) (BY SIMILARITY).
FT DISULFID 112 BY SIMILARITY.
FT DISULFID 178 BY SIMILARITY.
FT DISULFID 301 BY SIMILARITY.
FT DISULFID 380 443 BY SIMILARITY.
FT DISULFID 393 412 BY SIMILARITY.
FT DISULFID 402 477 BY SIMILARITY.
FT DISULFID 454 498 BY SIMILARITY.
FT DISULFID 503 553 BY SIMILARITY.
FT DISULFID 523 546 BY SIMILARITY.
FT CARBOHYD 287 287 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT LIPID 565 GPI-anchor amidated asparagine (By similarity).
SQ SEQUENCE 590 AA; 62950 MW; 0FB315D299659F58 CRC64;

Query Match 11.2%; Score 87; DB 1; Length 590;
Best Local Similarity 89.5%; Pred. No. 4.7;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 85 ADNYDQLVTRVVTHEMAHA 103
DB 238 ASRYDQLVTRVVTHEMAHA 256

RESULT 14
GP63 LEICH STANDARD; PRT; 599 AA.
AC P15706;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Leishmanolysin precursor (EC 3.4.24.36) (Cell surface protease)
DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface endopeptidase).
DE GP63.
OS Leishmania chagasi.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=44271;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92025976; PubMed=2320059;
RA Miller R.A., Reed S.G., Parsons M.;
RT "Leishmania gp63 molecule implicated in cellular adhesion lacks an Arg-Gly-Asp sequence.";
RL Mol. Biochem. Parasitol. 39:267-274(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92112918; PubMed=1370484;
RA Ramamoorthy R., Donelson J.E., Paetz K.E., Maybodi M., Roberts S.C., Wilson M.E.;
RT "Three distinct RNAs for the surface protease gp63 are differentially expressed during development of Leishmania donovani chagasi promastigotes to an infectious form.";
RL J. Biol. Chem. 267:1888-1895(1992).
CC -!- FUNCTION: Has an integral role during the infection of macrophages in the mammalian host.
CC -!- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and P1' and basic residues at P2 and P3'. A model nonapeptide is cleaved at -Ala-Tyr-Ileu-Lys-Iys-.
CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -!- SIMILARITY: Belongs to peptidase family M8.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
EMBL; M80672; AAA29238.1; -.

```

```

DR EMBL; M88527; AAA29235.1; -.
DR PIR; A44951; A44951.
DR KSSP; P08148; ILML.
DR MEROPS; M08.001; -.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR InterPro; IPR001577; Peptidase_M8.
DR Pfam; PF01457; Peptidase_M8.1.
DR PRINTS; PR00782; LSHMANOLYSIN.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydrolase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;
KW Zymogen; Signal; Cell adhesion; GPI-anchor; Lipoprotein.
KW SIGNAL 1 39 POTENTIAL.
FT PROPEP 40 97 ACTIVATION PEPTIDE.
FT CHAIN 98 574 LEISHMANOLYSIN.
FT PROPEP 575 599 REMOVED IN MATURE FORM (BY SIMILARITY).
FT METAL 261 261 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 262 262 BY SIMILARITY.
FT METAL 265 265 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 331 331 ZINC (CATALYTIC) (BY SIMILARITY).
FT DISULFID 122 139 BY SIMILARITY.
FT DISULFID 188 227 BY SIMILARITY.
FT DISULFID 311 383 BY SIMILARITY.
FT DISULFID 390 452 BY SIMILARITY.
FT DISULFID 403 422 BY SIMILARITY.
FT DISULFID 412 486 BY SIMILARITY.
FT DISULFID 463 507 BY SIMILARITY.
FT DISULFID 512 562 BY SIMILARITY.
FT DISULFID 532 555 BY SIMILARITY.
FT CARBOHYD 297 297 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 394 394 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT LIPID 574 574 GPI-anchor amidated asparagine (By similarity).
SQ SEQUENCE 599 AA; 63848 MW; 746730AE8E2A2E7C CRC64;

Query Match 11.2%; Score 87; DB 1; Length 599;
Best Local Similarity 89.5%; Pred. No. 4.8;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 85 ADNYDQLVTRVVTHEMAHA 103
DB 248 ASRYDQLVTRVVTHEMAHA 266

RESULT 15
GP63 LEIMA STANDARD; PRT; 602 AA.
AC P08148; P15906;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Leishmanolysin precursor (EC 3.4.24.36) (Cell surface protease)
DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface endopeptidase).
DE GP63.
OS Leishmania major.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A. AND SEQUENCE OF 101-123.
RX MEDLINE=88154764; PubMed=3346625;
RA Burton L.L., McMaster W.R.;
RT "Molecular cloning of the major surface antigen of leishmania.";
RL J. Exp. Med. 167:724-729(1988).
RN [2]
RP REVISIONS.
RA Burton L.L., McMaster W.R.;
RL J. Exp. Med. 171:589-589(1990).
RN [3]
RP GPI-ANCHOR.
RX MEDLINE=91009116; PubMed=2145267;
RA Schneider P., Ferguson M.A.J., McConville M.J., Mehler A., Honans S.W., Bordier C.;
RA "Structure of the glycosyl-phosphatidylinositol membrane anchor of

```


Query Match 11.2%; Score 87; DB 1; Length 602;
Best Local Similarity 89.5%; Pred. No. 4.8;
Matches 17; Conservative 0; Mismatches 2; Indels

Sequence	Matches	Mismatches	Indels	Gaps
Conservative	17	0	2	0
Radical	17	0	2	0

Search completed: August 2, 2004, 14:49:31
Job time : 6.3 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:39:17 ; Search time 29.7 Seconds
(without alignments)
1604.150 Million cell updates/sec

Title: US-09-543-407-22
Perfect score: 776
Sequence: 1 MKLKVAFAAIVVSGSALA.....DSSVMVRQVFGNNATANQY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL 25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_plage:*
- 10: sp_plant:*
- 11: sp_todent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES						
Result No.	Score	Query Match %		Length	DB ID	Description
		Match	%			
1	673	86.7	152	2	O33802	Q33802 salmonella
2	566.5	73.0	150	2	Q7X243	Q7X243 citrobacter
3	538	69.3	149	2	Q7X240	Q7X240 citrobacter
4	496.5	64.0	152	16	Q8CW63	Q8CW63 escherichia
5	421.5	54.3	150	2	Q7X237	Q7X237 enterobacte
6	305.5	39.4	76	2	Q54069	Q54069 salmonella
7	122	15.7	29	2	Q983J5	Q983J5 escherichia
8	115	14.8	139	16	Q8EIH3	Q8EIH3 shewanella
9	110	14.2	130	16	Q8EJI4	Q8EJI4 bradyrhizob
10	109.5	14.1	502	16	Q8EIH4	Q8EIH4 shewanella
11	109	14.0	1748	5	Q94821	Q94821 tetrahymena
12	109	14.0	2174	16	Q92JU8	Q92JU8 rhizobium n
13	106.5	13.7	3501	16	Q8Y106	Q8Y106 ralstonia s
14	106.5	13.7	3552	16	Q8XSD6	Q8XSD6 ralstonia s
15	106	13.7	179	2	O33801	O33801 salmonella
16	105	13.5	1422	16	Q8EFU3	Q8EFU3 shewanella

17	104	13.4	714	16	Q7U5X6	Q7u5x6 synechococ
18	102	13.1	191	10	Q7XDR3	Q7xdr3 oryza sativ
19	102	13.1	586	5	Q9NAJ4	Q9naj4 caenorhabdi
20	98	12.6	614	16	Q7TYR8	Q7tyr8 mycobacteri
21	98	12.6	615	16	P95249	P95249 mycobacteri
22	96.5	12.4	171	16	Q89J13	Q89j13 bradyrhizob
23	96	12.4	2310	5	Q9W2U7	Q9w2u7 drosophila
24	95.5	12.3	3859	16	Q98LNE	Q98ln6 rhizobium 1
25	95	12.2	589	16	Q7TXS3	Q7txs3 mycobacteri
26	95	12.2	590	16	O53309	O53309 mycobacteri
27	95	12.2	603	16	Q8VJ66	Q8vj66 mycobacteri
28	94.5	12.2	157	16	Q88HG0	Q88hg0 pseudomonas
29	94.5	12.2	738	5	O02402	O02402 pinctada fu
30	94.5	12.2	716	16	Q7UWZ8	Q7uwz8 rhodospirell
31	94	12.1	407	5	Q20151	Q20151 caenorhabdi
32	93.5	12.0	145	16	Q8U6N9	Q8u6n9 agrobacteri
33	93.5	12.0	645	16	Q7U1C5	Q7u1c5 mycobacteri
34	93.5	12.0	646	16	O53818	O53818 mycobacteri
35	93.5	12.0	1286	2	Q841Y5	Q841y5 campylobact
36	93.5	12.0	1333	16	Q8PD38	Q8pd38 xanthomonas
37	92.5	11.9	151	16	Q7UC21	Q7uc21 shigella fl
38	92.5	11.9	154	16	Q89J15	Q89j15 bradyrhizob
39	92.5	11.9	160	16	Q8CW64	Q8cw64 escherichia
40	92.5	11.9	160	16	Q83RU7	Q83r17 shigella fl
41	92.5	11.9	453	5	Q9NGF7	Q9ngf7 drosophila
42	92.5	11.9	453	5	Q9NGM8	Q9ngm8 drosophila
43	92	11.9	196	10	O22638	O22638 zea mays (m
44	91.5	11.8	151	2	Q7X238	Q7x238 enterobacte
45	91.5	11.8	453	5	Q9NGF6	Q9ngf6 drosophila

ALIGNMENTS

RESULT 1

O33802	PRELIMINARY;	PRT;	152 AA.
ID	O33802		
AC	O33802;		
DT	01-JAN-1998 (TRENBLrel. 05, Created)		
DT	01-JAN-1998 (TRENBLrel. 05, Last sequence update)		
DT	01-DEC-2001 (TRENBLrel. 19, Last annotation update)		
DE	Agfa protein (Fragment).		
GN	AGFA.		
OS	Salmonella typhimurium.		
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;		
OC	Enterobacteriaceae; Salmonella.		
OX	NCBI TaxID=602;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=98053981; PubMed=9393832;		
RA	Sukupolvi S.S., Lorentz R.G., Gordon J.I., Bian Z., Pfeifer J.D.,		
RA	Normark S.J., Rhen M.;		
RT	"Expression of thin, aggregative fimbriae promotes interaction of		
RT	Salmonella typhimurium SR-11 with mouse small intestinal epithelial		
RT	cells.";		
RL	Infect. Immun. 65:5320-5325 (1997).		
DR	EMBL; AJ000514; CAA04151.1; -.		
FT	NON TER	152	152
SQ	SEQUENCE	152 AA; 15401 MW; 9DA7DADC2364B006 CRC64;	
	Query Match	86.7%; Score 673; DB 2; Length 152;	
	Best Local Similarity	89.4%; Pred. No. 3.7e-47;	
	Matches 135; Conservative 3; Mismatches 13; Indels 0; Gaps 0;		
Qy	1	MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDPSTLSIYQVGSANAALALQ	60
Db	1	MKLLKVAFAAIVVSGSAVAGVVPQWGGGNGHNGSGPDPSTLSIYQVGSANAALALQ	60
Qy	61	SDARKSETTITQSGYNGADVGQADNDVQLVTVVTHEMAHADQWNAKSDITVGYGG	120
Db	61	SDARKSETTITQSGYNGADVGQADNDVQLVTVVTHEMAHADQWNAKSDITVGYGG	120
Qy	121	NNAALVNOTADSSVMVRQVFGNNATANQY	151

```

Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
|||||
RESULT 2
Q7X243 ID Q7X243 PRELIMINARY; PRT; 150 AA.
AC Q7X243;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Curlin-csgA protein.
GN CSGA.
OS Citrobacter sp. Fec2.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Citrobacter.
OX NCBI_TaxID=213763;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fec2;
RA Zogaj X., Bokranz W., Nimitz M., Romling U.;
RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RT Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158(2003).
DR EMBL; AJ515700; CAD56672.1; -.
SQ SEQUENCE 150 AA; 15016 MW; 1D7141B8D6973DC6 CRC64;

Query Match 73.0%; Score 566.5; DB 2; Length 150;
Best Local Similarity 77.5%; Pred. No. 1.6e-38;
Matches 117; Conservative 11; Mismatches 22; Indels 1; Gaps 1;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQW-GGGGNGHNGGSGPDSTLSIYQYGSANAALALQ 60
|||
Db 1 MKLLQVAFAAIVVSGSALAGVVPQWGGGGG-GGGGSSGSPSTLSIYQYGVNNAALALQ 59
|||
QY 61 SDARKSETTITQSGYNGADYQGGADNYDQLVTRVVTHEMAHADQWNAKNSDITVGYGG 120
|||||
Db 60 SDARKSDTHIHQHGNGADYQGGSDNSTIDITQFGKXNATIDQWNGKNSDITVSYGG 119
|||||
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
|||||
Db 120 HNAALVNQTASDSSVLRVQVGFNNATANQY 150
|||||

RESULT 3
Q7X240 ID Q7X240 PRELIMINARY; PRT; 149 AA.
AC Q7X240;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Curlin-csgA protein.
GN CSGA.
OS Citrobacter freundii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Citrobacter.
OX NCBI_TaxID=545;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fec4;
RA Zogaj X., Bokranz W., Nimitz M., Romling U.;
RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RT Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158(2003).
DR EMBL; AJ515701; CAD56672.1; -.
SQ SEQUENCE 149 AA; 15280 MW; 946DD52017F648FD CRC64;

Query Match 69.3%; Score 538; DB 2; Length 149;
Best Local Similarity 72.2%; Pred. No. 3.2e-36;
Matches 109; Conservative 17; Mismatches 23; Indels 2; Gaps 1;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSGPDSTLSIYQYGSANAALALQ 60
|||||

```

```

Db 1 MKLLKVAFAAIVVSGSALAGVVPQW--GGNHGCGSGNYGPDSSLSIYQYGSNNANALQ 58
61 SDARKSETTITQSGYNGADYQGGADNYDQLVTRVVTHEMAHADQWNAKNSDITVGYGG 120
|||||
Db 59 SDARKSDVTITQHGNGGAVVVGQGADDSITSLKCTGFQNSATIDQWNAKADISVITQFGG 118
|||||
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
|||||
Db 119 RNgALVNQTASDSNVLIQQVQVGFNNATANQH 149
|||||
RESULT 4
Q8CW63 ID Q8CW63 PRELIMINARY; PRT; 152 AA.
AC Q8CW63;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Major curlin subunit precursor.
GN CSGA OR C1306.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
DR EMBL; AE016759; AAN79779.1; -.
KW Complete proteome.
SQ SEQUENCE 152 AA; 15064 MW; 3BA57F34C1240E83 CRC64;

Query Match 64.0%; Score 496.5; DB 16; Length 152;
Best Local Similarity 67.8%; Pred. No. 7.7e-33;
Matches 103; Conservative 17; Mismatches 31; Indels 1; Gaps 1;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQW-GGGGNGHNGGSGPDSTLSIYQYGSANAALAL 59
|||||
Db 1 MKLLKVAFAAIVVSGSALAGVVPQVGGGNGHGGGNSGNSLNIYQYGGNSALAQ 60
|||||
QY 60 QSDARKSETTITQSGYNGADYQGGADNYDQLVTRVVTHEMAHADQWNAKNSDITVGYGG 119
|||||
Db 61 QADARNSDLTITQHGNGGADYQGGSDSSIDLTQRFQNSATLDQWNGKDSMTVTKQFG 120
|||||
QY 120 GNNALVNQTASDSSVMVRQVGFNNATANQY 151
|||||
Db 121 GGNGRAVDQTASNSSVNVTVQVGFNNATAHQY 152
|||||
RESULT 5
Q7X237 ID Q7X237 PRELIMINARY; PRT; 150 AA.
AC Q7X237;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Curlin-csgA protein.
GN CSGA.
OS Enterobacter sakazakii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Enterobacter.
OX NCBI_TaxID=28141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fec39;
RA Zogaj X., Bokranz W., Nimitz M., Romling U.;

```

```

RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RL Enterobacteriaceae isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4159(2003).
DR EMBL; AJ515702; CAD56678.1; -.
SQ SEQUENCE 150 AA; 15112 MW; 508BB2D872DF15F3 CRC64;

Query Match 54.3%; Score 421.5; DB 2; Length 150;
Best Local Similarity 58.9%; Pred. No. 9.2e-27;
Matches 89; Conservative 25; Mismatches 36; Indels 1; Gaps 1;

QY 1 MKLLKVAFAAIVVSGSALAGVQWGGGNNHGGSSGPDSTLSIYQYGSANAALALQ 60
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MKFIVKAAALAAIVVSGSAAVAGMIHQ-GGCHGCHGGVGGPSTLNLYQGGNSALALQ 59
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 61 SDARKSETTITQSGYGNAGDVQCADNYDQIVTRVVTHEMAHADQWNAKSDIVVGYGG 120
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 60 TDARNSVLNISTQGGNGADYGVGGSDSSINLTQNGFNSATLDQNSKDSWMNVSYGG 119
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 121 NNAALVNQATDSSVMVRQVFGNNATANQY 151
Db 120 LNALVDTQASNSIVVTVTQIGFNGEATAHQY 150
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 6
Q54069 PRELIMINARY; PRT; 76 AA.
AC Q54069;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DE SEF17 fimbria (Fragment).
GN AGPA.
OS Salmonella enteritidis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=592;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SE30;
RA Cox J.M., Eglezos S., Woolcock J.B.;
RT "Virulence of Salmonella enteritidis in chickens correlates with
RL colony morphology and expression of SEF17 fimbriae.";
RL Submitted (Apr-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U53207; AAA98671.1; -.
FT NON TER 1
FT NON TER 76
SQ SEQUENCE 76 AA; 7704 MW; 2FD5411241A7BCB1 CRC64;

Query Match 39.4%; Score 305.5; DB 2; Length 76;
Best Local Similarity 54.5%; Pred. No. 1e-17;
Matches 66; Conservative 4; Mismatches 6; Indels 45; Gaps 1;

QY 30 GNHNGGSSGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYGNAGDVQCADNYD 89
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 GNHNGGSSGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYGNAGDVQCADN-- 58
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 90 QIVTRVVTHEMAHADQWNAKSDITVQYVGGNNAALVNQATDSSVMVRQVFGNNATAN 149
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 59 -----STIELTONGFRNATID 75

QY 150 Q 150
Db 76 Q 76

RESULT 7
Q9S3J5 PRELIMINARY; PRT; 29 AA.
AC Q9S3J5;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAY-2001 (TREMBlrel. 16, Last annotation update)
DE Curlin subunit monomer (Fragment).

```

```

GN CSGA.
OC Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC TRANSPOSON-Insertion sequence IS1;
RC MEDLINE=99314153; PubMed=10386375;
RX La Regione R.M., Collighan R.J., Woodward M.J.;
RT "Non-curling of Escherichia coli O78:H80 isolates associated with
RT IS1 inserti on in csbg and reduced persistence in poultry infection.";
RL FEMS Microbiol. Lett. 175:247-253(1999).
DR EMBL; AJ131756; CAB45380.1; -.
FT NON TER 29
FT NON TER 29
SQ SEQUENCE 29 AA; 2789 MW; E290DFC07ABBB243 CRC64;

Query Match 15.7%; Score 122; DB 2; Length 29;
Best Local Similarity 89.7%; Pred. No. 0.0025;
Matches 26; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGG 29
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MKLLKVAATAAIVVSGSALAGVVPQYGGG 29
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 8
Q8EIH3 PRELIMINARY; PRT; 139 AA.
AC Q8EIH3;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Minor curlin subunit CsgB, putative.
GN S00866.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.
OX NCBI_TaxID=70863;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WR-1;
RX MEDLINE=22237686; PubMed=12368813;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eilen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
RA Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,
RA Mueller J., Khouri H., Gill J., Utterback T.R., McDonald L.A.,
RA Feldblyum T.V., Smith H.O., Venter J.C., Neilson K.H., Fraser C.M.;
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis.";
RL Nat. Biotechnol. 20:1118-1123(2002).
DR EMBL; AE015532; AAN53942.1; -.
DR TIGR; S00866; -.
RW Complete proteome.
SQ SEQUENCE 139 AA; 14811 MW; 41E1CFA76957920 CRC64;

Query Match 14.8%; Score 115; DB 16; Length 139;
Best Local Similarity 28.3%; Pred. No. 0.061;
Matches 30; Conservative 19; Mismatches 33; Indels 24; Gaps 2;

QY 39 SGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYGNAGDVQCADNYDQIVTRVTH 98
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 41 SGRDNLIDIVQQTANGGI-----VFQSGSDNSAYVTQAGDNISLVLTQIGT- 87
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 99 EMAHADQWNAKSDITVQYVGGNNAALVNQATDSSVMVRQVFGN 144
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 88 -----NNEVQLLVQGAQNKASITQIGNDNLVQLNLQSGN 122
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 9

```

```

Q89J14
ID Q89J14 PRELIMINARY; PRT; 130 AA.
AC Q89J14;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE BIL5299 protein.
GN BIL5299.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
CX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RX MEDLINE=22484998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA Sasamoto S., Watanabe A., Iidesawa K., Iriguchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimpo S., Tsuruoka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110.";
RL DNA Res. 9:189-197(2002).
DR EMBL; AP005954; BAC50564.1; -.
KW Complete proteome.
SQ SEQUENCE 130 AA; 12699 MW; ACFB2D66A48D260F CRC64;

Query Match 14.2%; Score 110; DB 16; Length 130;
Best Local Similarity 25.8%; Pred. No. 0.14;
Matches 39; Conservative 28; Mismatches 56; Indels 28; Gaps 5;

QY 4 LKVAFAAIVVSGSALAGVVPQGGNGHNGSGSPDSTLSIYQVGSANAALALQSDA 63
Db 1 MRITVLVATAIALSALTVDAA-----AGNSA-----SVLPFGITNSSFISQIGS 45

QY 64 RKSETITQSGYNGADVQGGADNYDQLVTRVTHEMAHADQW-----NAKNSDITVQY 119
Db 46 TSNNAITLQFGATNTATTTLQTS-----LLTVNTAVTGGGTATASNTALT-CGVG 96

QY 120 GNNALVNQTASDSSVMVRQVGFNNATANO 150
Db 97 GNSSLIGQIGANTAGVQGLNGSTILQ 127

RESULT 10
Q8EIH4
ID Q8EIH4 PRELIMINARY; PRT; 502 AA.
AC Q8EIH4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Conserved hypothetical protein.
GN S00865.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.
CX NCBI_TaxID=70863;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NR-1;
RX MEDLINE=22297686; PubMed=12368813;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
RA Vamathevan J., Weidman J., Imprim M., Lee K., Berry K., Lee C.,
RA Mueller J., Khouri H., Gill J., Utterback T.R., McDonald L.A.,
RA Feldblyum T.V., Smith H.O., Venter J.C., Nealon K.H., Fraser C.M.;
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis.";
RL Nat. Biotechnol. 20:1118-1123(2002).
DR EMBL; AE015532; AAN53941.1; -.

DR TIGR; S00865; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 502 AA; 52441 MW; D08CA23D6C46B62D CRC64;

Query Match 14.1%; Score 109.5; DB 16; Length 502;
Best Local Similarity 23.8%; Pred. No. 0.79; 50; Indels 49; Gaps 5;
Matches 39; Conservative 26; Mismatches 26;

QY 29 GGNHNGGSGSPDSTLSIYQVGSANAA-----LALQSDARKSETTITQSGYNG 78
Db 347 GDNNELVAFATGDSNTEISQEGDANFAYVDATGNDNEVNVQDQDNETIIVG-NNN 405

QY 79 ADVG-----QGADNYQLVTRVTHEMAHADQW-----N 107
Db 406 ADVTALQHRGDLNLITIEGDENAABIT-----QAGSGNVVGGDSFSAASFQVS 458

QY 108 AKNSDITVQYGGNNAALVNQTASDSSVMVRQVGFNNATANOY 151
Db 459 GDNNSLMITQGNLNLVLSQAGNNNSISVTQSGDMNVATVQY 502

RESULT 11
Q94821
ID Q94821 PRELIMINARY; PRT; 1748 AA.
AC Q94821; P92146; P92145; P92144; P92143; P92142; P92141; Q94820;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE CNJB protein.
GN CNJB.
OS Tetrahymena thermophila.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
OC Tetrahymenina; Tetrahymena.
CX NCBI_TaxID=5911;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88189811; PubMed=3357771;
RA Martindale D.W., Taylor F.M.;
RT "Multiple introns in a conjugation-specific gene from Tetrahymena
RT thermophila.";
RL Nucleic Acids Res. 16:2189-2201(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94051569; PubMed=8233798;
RA Taylor F.M., Martindale D.W.;
RT "Retroviral-type zinc fingers and glycine-rich repeats in a protein
RT encoded by cnjB, a Tetrahymena gene active during meiosis.";
RL Nucleic Acids Res. 21:4610-4614(1993).
DR EMBL; X06462; CAB37323.1; -.
DR PIR; S42136; S42136.
DR HSP; P05888; IAAF.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF00098; zf_CCHC; 7.
DR PRINTS; PR00939; C2HCZNFINGER.
DR SMART; SM00343; Znf_C2HC; 7.
DR PROSITE; PS0158; ZF_CCHC; 7.
DR CONFLICT 251 251 M -> I (IN REF. 1).
DR CONFLICT 256 256 I -> N (IN REF. 1).
SQ SEQUENCE 1748 AA; 199624 MW; 0B03F210104008A3 CRC64;

Query Match 14.0%; Score 109; DB 5; Length 1748;
Best Local Similarity 33.6%; Pred. No. 3.9;
Matches 37; Conservative 11; Mismatches 32; Indels 30; Gaps 6;

QY 25 QWGGGNGHNGG---GNSSGPDSTLSIYQVGSANAALALQSDARKSETTIT---OSGYGN 77
Db 1640 QFSGGNGSGGSGWTSSGSDWN-----QGSNVQESTTSSGGWGGSGGN 1685

QY 78 GADVGGGA-DNYDQLVTRVTHEMAHADQWNAKNSDITVQY---YGNNA 124
Db 1686 QTGGGWSNDNQOQ-----QNETGGGGGSGNSQNTNNESSWGSNNQA 1729

```



```

RX MEDLINE=21691879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Arlat M., Billault A., Brothier P., Camus J.C., Cattolico L.,
RA Chandler M., Choisme N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Sigulier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
DR EMBL; AL646061; CAD14589.1; -.
DR GO; GO:0004519; F:endonuclease activity; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001604; Endonuclease.
DR InterPro; IPR008619; Fil haemagg.
DR InterPro; IPR008638; Haemagg act.
DR Pfam; PF05594; Fil haemagg; 20.
DR Pfam; PF05860; Haemagg act; 1.
DR PROSITE; PS01070; NUCLEASE_NON_SPSC; 1.
DR Complete proteome.
KW SEQUENCE 3501 AA; 348421 MW; 290841C99018A107 CRC64;
SQ
Query Match 13.7%; Score 106.5; DB 16; Length 3501;
Best Local Similarity 29.5%; Pred. No. 14;
Matches 43; Conservative 19; Mismatches 45; Indels 39; Gaps 7;
QY 15 SCSALAGVVPQWGGGNNHGGG-NSSGPDSTLSIYQGSANAA-----56
Db 2431 SCSHFSTAGFSWGDGRNVGGGPNSSG-----VGLAPYGAHSAADNAAGNSSQNASVWIG 2486
QY 57 LALOSDARKSETTITQSGYNGADVG---QCADNVQDLVTRVVTHEMA--HADOWNAKN 110
Db 2487 KSVQVQARIGDITVSGSGISALSDVDLLAKQK-----VDIVAGNDSRSHED-----H 2535
QY 111 SDITVQYGGNNALVNOTASDSSVM 136
Db 2536 SDRTIGDLGGYSGTVGVSASSTL 2561
RESULT 14
Q8XSD6 PRELIMINARY; PRT; 3552 AA.
ID Q8XSD6 AC Q8XSD6 PRELIMINARY; PRT; 3552 AA.
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Probable hemagglutinin-related protein.
GN RSP0540.0R RS06117.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
OC NCBI_TaxID=305;
RP [1]
RN SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RX MEDLINE=21691879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Arlat M., Billault A., Brothier P., Camus J.C., Cattolico L.,
RA Chandler M., Choisme N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Sigulier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
DR EMBL; AL646079; CAD17691.1; -.
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
DR GO; GO:0004519; F:endonuclease activity; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001604; Endonuclease.
DR InterPro; IPR008619; Fil haemagg.
DR InterPro; IPR008638; Haemagg act.
DR Pfam; PF05594; Fil haemagg; 20.
DR Pfam; PF05860; Haemagg act; 1.
DR Complete proteome.
KW SEQUENCE 3501 AA; 348421 MW; 290841C99018A107 CRC64;
SQ
Query Match 14.0%; Score 109; DB 16; Length 2174;
Best Local Similarity 27.0%; Pred. No. 5.1;
Matches 40; Conservative 20; Mismatches 52; Indels 36; Gaps 7;
QY 11 AIVVSGSALAGVVPQ--WGGGNNHGGGNSGPDSTLSIYQGS-----ANAA-----56
Db 693 AIAATAGAGAVGLAQSIGGGGN---GCGATGGDAGFGSFQGGGGGGGVANTANVGFK 749
QY 57 -LALOSDARKSETTITQSGYNGADVGQADNVQDLVTRVVTHEMAHADOWNAKNSDITV 115
Db 750 GLTLTQGSAAAGIVAQS-VGGGGGTGTGTSYSAGI-----GFTASVAV 793
QY 116 GQYGGNAA--LVNOTASDSSVWVRQVG 141
Db 794 GGTGGNGAGGVEVSVLDSARTQGG 821
RESULT 13
Q8Y106 PRELIMINARY; PRT; 3501 AA.
ID Q8Y106 AC Q8Y106 PRELIMINARY; PRT; 3501 AA.
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Probable hemagglutinin-related protein.
GN RSC0887.0R RS06116.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
OC NCBI_TaxID=305;
RP [1]
RN SEQUENCE FROM N.A.
RC STRAIN=GM11000;
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:35:42 ; Search time 44.9 Seconds
(without alignments)
950.215 Million cell updates/sec

Title: US-09-543-407-24
Perfect score: 775
Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMRVQVFGNNATANQY 151

Scoring table: BLOSUM62

Searched: 1586107 seqs, 282547505 residues
Gapop 10.0 , Gapext 0.5

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	775	100.0	151	3	AAB36352
2	714	92.1	151	3	AAB36347
3	700	90.3	151	3	AAB36355
4	690	89.0	151	2	AAR74625
5	690	89.0	151	3	AAB36341
6	689	88.9	151	2	AAW23570
7	655	84.5	151	3	AAB36346
8	609	78.6	151	3	AAB36353
9	609	78.6	151	3	AAB36351
10	608	78.5	151	3	AAB36349
11	606	78.2	151	3	AAB36350
12	600	77.4	151	3	AAB36354
13	575	74.2	151	3	AAB36348
14	533	68.8	151	3	AAB36343
15	528	68.1	151	7	ABR82651
16	505	65.2	120	2	AAW23569
17	505	65.2	120	2	AAW23569
18	455	58.7	142	2	AAW23564
19	383	49.4	122	2	AAW23563
20	237	30.6	45	3	AAW23561
21	132	17.0	22	3	AAW23561
22	123	15.9	23	3	AAW23561
23	123	15.9	23	3	AAW23561
24	123	15.9	23	3	AAW23561
25	115	14.8	22	3	AAW23561

26	115	14.8	22	3	AAB36339
27	115	14.8	22	3	AAB36320
28	113	14.6	24	7	ABR82644
29	109	14.1	23	3	AAB36340
30	109	14.1	23	3	AAB36324
31	109	14.1	23	3	AAB36319
32	98.5	12.7	151	3	AAB36342
33	98	12.6	26	7	ABR82649
34	98	12.6	26	7	ABR82645
35	98	12.6	903	2	AAW35006
36	96.5	12.5	151	3	AAB36344
37	96	12.4	19	3	AAB36323
38	96	12.4	19	3	AAB36336
39	96	12.4	19	3	AAB36328
40	92.5	11.9	850	4	ABW5764
41	92.5	11.9	1028	4	ABW5764
42	92	11.9	23	3	AAB36331
43	91	11.7	688	5	ABP74039
44	90.5	11.7	378	4	ABW5764
45	90.5	11.7	520	6	AAO16497

ALIGNMENTS

RESULT 1
AAB36352
ID AAB36352 standard; protein; 151 AA.
XX
AC AAB36352;
XX
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa::PT3#7 amino acid sequence SEQ ID NO:24.
XX
XX Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
XX vaccine; immune response; immunogen.
XX
XX Salmonella enteritidis.
XX Escherichia coli.
XX Synthetic.
XX
XX WO2000060102-A2.
XX
XX 12-OCT-2000.
XX
XX 05-APR-2000; 2000WO-CA000356.
XX
XX 05-APR-1999; 99US-0127888P.
XX
XX (UYVI-) UNIV VICTORIA.
XX
XX White AP, Doran JL, Collison SK, Kay WW;
XX
XX WPT; 2000-672631/65.
XX
XX N-PSDB; AAC64628.
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
XX which encodes foreign epitope or antigen, expresses recombinant Agfa
XX protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 138; 139pp; English.

The present invention describes a recombinant agfa gene (1) where a segment of the gene has been replaced by a segment of a foreign DNA sequence which encodes a foreign epitope or antigen. Also described are: (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation dependent assembly system of strains of Salmonella, Escherichia coli and Enterobacteriaceae for the production of fimbriae comprising recombinant Agfa, CsgA and Agfa-homologue fimbrin subunits, respectively; (2) directing recombination of a recombinant gene into the chromosome of the homologous species; (3) directing recombination of a recombinant gene back into the chromosome of the homologous species, replacing the native

CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant AgfA
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fibrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;

Query Match 100.0%; Score 775; DB 3; Length 151;
 Best Local Similarity 100.0%; Pred. No. 1.6e-68;
 Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQYGSANAALALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQYGSANAALALQ 60
 QY 61 SDARKSETTITQSGYNGADVGQGDNSTIETQNGFRNNATIDQWNAKNDQLVTRVVT 120
 DB 61 SDARKSETTITQSGYNGADVGQGDNSTIETQNGFRNNATIDQWNAKNDQLVTRVVT 120
 QY 121 HEMAHANQATSDSSVMVRQVGFNNATANQY 151
 DB 121 HEMAHANQATSDSSVMVRQVGFNNATANQY 151

RESULT 2
 AAB36347
 ID AAB36347 standard; protein; 151 AA.
 AC AAB36347;
 XX 26-FEB-2001 (first entry)
 DE AgfA::PT3#2 amino acid sequence SEQ ID NO:14.
 KW Salmonella; agfA; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 XX Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX WO200060102-A2.
 XX 12-OCT-2000.
 XX 05-APR-2000; 2000WO-CA000356.
 XX 05-APR-1999; 99US-0127888P.
 XX (UYVI-) UNIV VICTORIA.
 XX White AP, Doran JL, Collison SK, Kay WW;
 XX WPI: 2000-672631/65.
 XX N-PSDB; AAC64623.
 XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence
 XX which encodes foreign epitope or antigen, expresses recombinant AgfA
 XX protein useful for eliciting immune response in animal.
 PS Disclosure; Page 136; 139pp; English.

XX The present invention describes a recombinant agfA gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/7AF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC AgfA, CsaA and AgfA-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant AgfA
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fibrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;

Query Match 92.1%; Score 714; DB 3; Length 151;
 Best Local Similarity 91.1%; Pred. No. 1.8e-62;
 Matches 144; Conservative 0; Mismatches 0; Indels 14; Gaps 2;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQYGSANAALALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQYGSANAALALQ 60
 QY 61 SDARKSETTITQSGYNGADVGQGDNSTIETQNGFRNNATIDQWNAKNDQLVTRVVT 113
 DB 61 SDARKSETTITQSGYNGADVGQGDNSTIETQNGFRNNATIDQWNAKNDQLVTRVVT 120
 QY 114 LVTRVVTHEMAHANQATSDSSVMVRQVGFNNATANQY 151
 DB 121 LVTRVVTHEMAHA-----SVNVRQVGFNNATANQY 151

RESULT 3
 AAB36355
 ID AAB36355 standard; protein; 151 AA.
 AC AAB36355;
 XX 26-FEB-2001 (first entry)
 DE AgfA::PT3#10 amino acid sequence SEQ ID NO:30.
 KW Salmonella; agfA; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 XX Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX WO200060102-A2.
 XX 12-OCT-2000.
 XX 05-APR-2000; 2000WO-CA000356.
 XX 05-APR-1999; 99US-0127888P.
 XX (UYVI-) UNIV VICTORIA.

XX White AP, Doran JL, Collison SK, Kay WW;
 XX WPI; 2000-672631/65.
 DR N-PSDB; AAC64631.
 XX
 XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant AgfA
 PT protein useful for eliciting immune response in animal.
 XX
 XX Disclosure; Page 139; 139pp; English.
 XX
 CC The present invention describes a recombinant agfA gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (Sf17/far) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant AgfA
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 XX Sequence 151 AA;
 Query Match 90.3%; Score 700; DB 3; Length 151;
 Best Local Similarity 89.9%; Pred. No. 4.2e-61;
 Matches 143; Conservative 0; Mismatches 0; Indels 16; Gaps 2;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
 QY 61 SDARKSETTITQSGYNGADVGQGDNSTIELTQNGFRNNATIDOWNAKNYDQLVTRVVT 120
 DB 61 SDARKSETTITQSGYNGADVGQGDNSTIELTQNGFRNNATIDOWNAKNYDQLVTRVVT 112
 QY 121 HEWAHA-----NCTASDSSVMVROVGFGNATANY 151
 DB 113 HEWAHAGNNAALVNQTSDDSSVMVROVGFGNATANY 151
 RESULT 4
 AAR74625
 ID AAR74625 standard; protein; 151 AA.
 XX AAR74625;
 XX
 XX 25-MAR-2003 (revised)
 DT 26-JUN-1995 (first entry)
 XX
 XX AgfA sequence.
 DE
 XX Salmonella; AgfA; vaccine.
 XX
 OS Salmonella.
 XX

PN WO9425598-A2.
 XX
 XX 10-NOV-1994.
 XX
 XX 26-APR-1994; 94WO-IB000207.
 PF
 XX 26-APR-1993; 93US-00054452.
 PR
 XX (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
 PA (KING/) KING J.
 XX
 XX Kay WW, Collinson SK, Clouthier SC, Doran JL;
 PI
 XX WPI; 1994-358275/44.
 XX
 XX N-PSDB; AAQ87467.
 DR
 XX
 XX Eliciting an immune response to Salmonella - using attenuated Salmonella
 PT strains, vector constructs, or compsns. contg. fimbrial type proteins.
 PT
 XX Disclosure; Fig 7B; 95pp; English.
 PS
 XX The Salmonella AgfA protein and DNA are used in vaccine and genetic
 CC immunization compositions, respectively, to elicit an immune response to
 CC Salmonella in animals (e.g. food producing animals) and humans. (Updated
 CC on 25-MAR-2003 to correct PN field.)
 CC
 XX Sequence 151 AA;
 SQ
 Query Match 89.0%; Score 690; DB 2; Length 151;
 Best Local Similarity 91.4%; Pred. No. 4.1e-60;
 Matches 138; Conservative 2; Mismatches 11; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
 QY 61 SDARKSETTITQSGYNGADVGQGDNSTIELTQNGFRNNATIDOWNAKNYDQLVTRVVT 120
 DB 61 SDARKSETTITQSGYNGADVGQGDNSTIELTQNGFRNNATIDOWNAKNSDITVQYGG 120
 QY 121 HEWAHANQTAASDSSVMVROVGFGNATANY 151
 DB 121 NNAALVNQTSDDSSVMVROVGFGNATANY 151
 RESULT 5
 AAB36341
 ID AAB36341 standard; protein; 151 AA.
 XX
 XX AAB36341;
 AC
 XX 26-FEB-2001 (first entry)
 DT
 XX
 XX Salmonella enteritidis AgfA amino acid sequence SEQ ID NO:5.
 DE
 XX Salmonella; agfA; chromosomal gene replacement; fimbrial; epitope;
 KW vaccine; immune response; immunogen.
 KW
 XX Salmonella enteritidis.
 OS
 XX WO2000060102-A2.
 PN
 XX 12-OCT-2000.
 PD
 XX
 XX 05-APR-2000; 2000WO-CA000356.
 PF
 XX 05-APR-1999; 99US-0127888P.
 PR
 XX (UYVI-) UNIV VICTORIA.
 PA
 XX White AP, Doran JL, Collison SK, Kay WW;
 PI
 XX WPI; 2000-672631/65.
 DR

DR N-PSDB; AAC64617.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 135; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC agfa, CsgA and Agfa-homologue fimbryn subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbryn protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;
 Query Match 89.0%; Score 690; DB 3; Length 151;
 Best Local Similarity 91.4%; Pred. No. 4.1e-60;
 Matches 138; Conservative 2; Mismatches 11; Indels 0; Gaps 0;
 QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGGNGSGDPSTLSIYQGSANAALALQ 60
 DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGGNGSGDPSTLSIYQGSANAALALQ 60
 QY 61 SPARKSETTITQSGYNGADVGQGDNSTIELTQGFRRNATIDQNAKNYDQLVTRVVT 120
 DB 61 SPARKSETTITQSGYNGADVGQGDNSTIELTQGFRRNATIDQNAKNYDQLVTRVVT 120
 QY 121 HEMAHANQTSADSSVMVRQVGFNNATANQY 151
 DB 121 NPAALVNTQTSADSSVMVRQVGFNNATANQY 151
 RESULT 6
 AAW23570
 ID AAW23570 standard; protein; 151 AA.
 XX
 AC AAW23570;
 XX
 DT 25-MAR-2003 (revised)
 DT 29-SEP-1997 (first entry)
 XX
 DE Salmonella enteritidis 27655-3b agfa.
 XX
 KW Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody.
 XX
 OS Salmonella enteritidis.
 XX
 PH Key Location/Qualifiers
 FT Misc-difference 123
 FT /note= "Encoded by GCC"
 XX

PN US5635617-A.
 XX
 PD 03-JUN-1997.
 XX
 PF 26-APR-1994; 94US-00233788.
 XX
 PR 26-APR-1993; 93US-00054452.
 XX
 XX (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
 FA
 XX Collinson SK, Kay WW, Doran JL;
 PI
 XX WPI; 1997-309886/28.
 DR
 XX N-PSDB; AAT74142.
 XX
 PT Isolated Salmonella gene agfa - used for diagnosis of Salmonella or
 PT enteropathogenic bacteria of the Enterobacteria family.
 XX
 PS Example 2; Fig 7; 85pp; English.
 XX
 CC The present sequence represents agfa encoded by the full agfa gene
 CC derived from Salmonella enteritidis 27655-3b. The nucleic acid can be
 CC used to provide diagnostic assays for Salmonella and/or enteropathogenic
 CC bacteria of the family Enterobacteria. It can also be used to provide
 CC proteins and antibodies which can be used for assays. The nucleic acid
 CC sequence can be used to provide probes or primers which can specifically
 CC hybridise to nucleic acid molecules from greater than 99% of Salmonella
 CC strains that are pathogenic to warm-blooded animals relative to nucleic
 CC acid molecules from virtually all other microbial organisms. (Updated on
 CC 25-MAR-2003 to correct PF field.)
 XX
 SQ Sequence 151 AA;
 Query Match 88.9%; Score 689; DB 2; Length 151;
 Best Local Similarity 91.4%; Pred. No. 5.2e-60;
 Matches 138; Conservative 2; Mismatches 11; Indels 0; Gaps 0;
 QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGGNGSGDPSTLSIYQGSANAALALQ 60
 DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGGNGSGDPSTLSIYQGSANAALALQ 60
 QY 61 SPARKSETTITQSGYNGADVGQGDNSTIELTQGFRRNATIDQNAKNYDQLVTRVVT 120
 DB 61 SPARKSETTITQSGYNGADVGQGDNSTIELTQGFRRNATIDQNAKNYDQLVTRVVT 120
 QY 121 HEMAHANQTSADSSVMVRQVGFNNATANQY 151
 DB 121 NPAALVNTQTSADSSVMVRQVGFNNATANQY 151
 RESULT 7
 AAB36346
 ID AAB36346 standard; protein; 151 AA.
 XX
 AC AAB36346;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Agfa::PT3#1 amino acid sequence SEQ ID NO:12.
 XX
 KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 PN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX

```
PR 05-APR-1999; 99US-0127888P.
XX (UUVI-) UNIV VICTORIA.
XX White AP, Doran JL, Collison SK, Kay WW;
XX WPI; 2000-672631/65.
XX N-PSDB; AAC64622.
XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX Disclosure; Page 135; 139pp; English.
XX The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (Sf17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX Sequence 151 AA;
XX
XX Query Match 84.5%; Score 655; DB 3; Length 151;
XX Best Local Similarity 79.8%; Pred. No. 1.2e-56;
XX Matches 134; Conservative 0; Mismatches 0; Indels 34; Gaps 2;
XX
XX QY 1 MKLLKVAAPAAIIVVSGSALAGVVPQGGGNGHNGGNSGDPSTLSIYQGSANAALALQ 60
XX DB 1 MKLLKVAAPAAIIVVSGSALAGVVPQGGGNGHNGGNSGDPSTLSIYQGSANAALALQ 60
XX
XX QY 61 SDARKSETTITQSGYNGADVGQGDADNSTIELTQNGFRNNATIDQWNAKSDITVGYGG 109
XX DB 61 SDARKSETTITQSGYNGADVGQGDADNSTIELTQNGFRNNATIDQWNAKSDITVGYGG 120
XX
XX QY 110 -----NYDOLVTRVTHEMAHANTQASDSSVMVVRQVFGNNATANQY 151
XX DB 121 NNAALVNYDQVTRVTHEMAHA-----NNATANQY 151
XX
XX RESULT 8
XX AAB36353 standard; protein; 151 AA.
XX AC AAB36353;
XX DT 26-FEB-2001 (first entry)
XX DE AgfA::PT3#8 amino acid sequence SEQ ID NO:26.
XX KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
XX vaccine; immune response; immunogen.
```

```
XX Salmonella enteritidis.
XX Escherichia coli.
XX Synthetic.
XX WO2000060102-A2.
XX 12-OCT-2000.
XX 05-APR-2000; 2000WO-CA0000356.
XX 05-APR-1999; 99US-0127888P.
XX (UUVI-) UNIV VICTORIA.
XX White AP, Doran JL, Collison SK, Kay WW;
XX WPI; 2000-672631/65.
XX N-PSDB; AAC64629.
XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX Disclosure; Page 138; 139pp; English.
XX The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (Sf17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX Sequence 151 AA;
XX
XX Query Match 78.6%; Score 609; DB 3; Length 151;
XX Best Local Similarity 82.1%; Pred. No. 4.2e-52;
XX Matches 124; Conservative 4; Mismatches 23; Indels 0; Gaps 0;
XX
XX QY 1 MKLLKVAAPAAIIVVSGSALAGVVPQGGGNGHNGGNSGDPSTLSIYQGSANAALALQ 60
XX DB 1 MKLLKVAAPAAIIVVSGSALAGVVPQGGGNGHNGGNSGDPSTLSIYQGSANAALALQ 60
XX
XX QY 61 SDARKSETTITQSGYNGADVGQGDADNSTIELTQNGFRNNATIDQWNAKSDITVGYGG 120
XX DB 61 LVTRVTHEMAHAGYNGADVGQGDADNSTIELTQNGFRNNATIDQWNAKSDITVGYGG 120
XX
XX QY 121 HEMAHAHQATASDSSVMVVRQVFGNNATANQY 151
XX DB 121 NNAALVNYDQVTRVTHEMAHA-----NNATANQY 151
XX
XX RESULT 9
```

AAB36351
 ID AAB36351 standard; protein; 151 AA.
 AC AAB36351;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Agfa::PT3#6 amino acid sequence SEQ ID NO:22.
 DE Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 PN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay WW;
 XX
 DR WPI; 2000-672631/65.
 DR N-PSDB; AAC64627.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 137; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;
 XX
 Query Match 78.6%; Score 609; DB 3; Length 151;
 Best Local Similarity 73.6%; Pred. No. 4.2e-52;
 Matches 128; Conservative 0; Mismatches 0; Indels 46; Gaps 2;
 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQVGSANAALQ 60
 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQVGSANAALQ 60

61 SDARKSETTITQSGYNGADVGCGADNNTIETQGFNNATIDQNAKVDQVTRVVT 120
 61 SDARKSETTITQSGYNGADVGCGADNNTIETQGFNNATIDQNAKVDQVTRVVT 97
 121 HEMAHA-----NQTSADSSVMVRQVGFNGNATANQY 151
 98 HEMAHADQWNAKNSDITVGQYGGNNAALVNQTSADSSVMVRQVGFNGNATANQY 151
 RESULT 10
 AAB36349
 ID AAB36349 standard; protein; 151 AA.
 AC AAB36349;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Agfa::PT3#4 amino acid sequence SEQ ID NO:18.
 DE Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 PN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay WW;
 XX
 DR WPI; 2000-672631/65.
 DR N-PSDB; AAC64625.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 136; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;
 XX
 Query Match 78.6%; Score 609; DB 3; Length 151;
 Best Local Similarity 73.6%; Pred. No. 4.2e-52;
 Matches 128; Conservative 0; Mismatches 0; Indels 46; Gaps 2;
 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQVGSANAALQ 60
 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQVGSANAALQ 60

CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;

Query Match 78.5%; Score 608; DB 3; Length 151;
 Best Local Similarity 82.1%; Pred. No. 5.2e-52;
 Matches 124; Conservative 5; Mismatches 22; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDPSTLSIYQYGSANAALALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDPSTLSIYQYGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGADVGQADNSTIETLTONGFRNNATIDOWNAKNYDQLVTRVVT 120
 DB 61 SDARKSETTITQSGYNGADVGQADNSTIETLTONGFRNNATIDOWNAKNSDITVGYGG 120

QY 121 HEMAHANQTSADSSVMVRQVGFNNATANQY 151
 DB 121 NNAALVNQTSADSSVMVRQVGFNNATANQY 151

RESULT 11
 AAB36350
 ID AAB36350 standard; protein; 151 AA.
 XX
 AC AAB36350;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Agfa::PT3#5 amino acid sequence SEQ ID NO:20.
 XX
 KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 PN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay WW;
 XX
 DR N-PSDB; AAC64626.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 137; 139pp; English.

The present invention describes a recombinant agfa gene (I) where a segment of the gene has been replaced by a segment of a foreign DNA sequence which encodes a foreign epitope or antigen. Also described are: (1) use of thin aggregative fimbriae (SEF17/1AF) nucleation depended assembly system of strains of Salmonella, Escherichia coli and Enterobacteriaceae for the production of fimbriae comprising recombinant Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2) directing recombination of a recombinant gene into the chromosome of the homologous species; (3) directing recombination of a recombinant gene back into the chromosome of the homologous species, replacing the native copy of that gene; and (4) eliciting an immune response in an animal, comprising separating an amino acid polymer comprising a recombinant Agfa protein containing a replacement segment or segments of foreign amino acid sequence or sequences grown on a Salmonella, E. coli or

CC Enterobacteriaceae host cell, from the host cell and introducing the polymer into the animal in conjunction with a carrier or diluent. (I) is useful for the expression of recombinant Agfa protein which is useful for eliciting an immune response in an animal. In a fimbrial presentation system the heterologous antigens are presented in high numbers (up to 500,000 copies/cell), the hybrid fimbrin protein possesses both the immunogenicity and adhesion properties relevant for an efficient live vaccine, the carrier fimbrial subunit proteins are usually strong immunogens, which may be important for directing an immune response against the inserted epitope, and hybrid fimbriae are easy and inexpensive to purify in large amount. The present sequence is given in the exemplification of the present invention
 XX
 SQ Sequence 151 AA;

Query Match 78.2%; Score 606; DB 3; Length 151;
 Best Local Similarity 82.1%; Pred. No. 8.2e-52;
 Matches 124; Conservative 5; Mismatches 22; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDPSTLSIYQYGSANAALALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDPSTLSIYQYGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGADVGQADNSTIETLTONGFRNNATIDOWNAKNYDQLVTRVVT 120
 DB 61 SDARKYDQLVTRVVTHEMAHAGCGADNSTIETLTONGFRNNATIDOWNAKNSDITVGYGG 120

QY 121 HEMAHANQTSADSSVMVRQVGFNNATANQY 151
 DB 121 NNAALVNQTSADSSVMVRQVGFNNATANQY 151

RESULT 12
 AAB36354
 ID AAB36354 standard; protein; 151 AA.
 XX
 AC AAB36354;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Agfa::PT3#9 amino acid sequence SEQ ID NO:28.
 XX
 KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 PN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay WW;
 XX
 DR N-PSDB; AAC64630.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 138; 139pp; English.

The present invention describes a recombinant agfa gene (I) where a segment of the gene has been replaced by a segment of a foreign DNA sequence which encodes a foreign epitope or antigen. Also described are: (1) use of thin aggregative fimbriae (SEF17/1AF) nucleation depended assembly system of strains of Salmonella, Escherichia coli and Enterobacteriaceae for the production of fimbriae comprising recombinant Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2) directing recombination of a recombinant gene into the chromosome of the homologous species; (3) directing recombination of a recombinant gene back into the chromosome of the homologous species, replacing the native copy of that gene; and (4) eliciting an immune response in an animal, comprising separating an amino acid polymer comprising a recombinant Agfa protein containing a replacement segment or segments of foreign amino acid sequence or sequences grown on a Salmonella, E. coli or

CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 XX Sequence 151 AA;
 Query Match 77.4%; Score 600; DB 3; Length 151;
 Best Local Similarity 81.5%; Pred. No. 3.2e-51;
 Matches 123; Conservative 3; Mismatches 25; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
 Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
 QY 61 SDARKSETTITQSGYNGADVGGADNSTIETQNGFRNNATIDQWNAKNDITVGVYGG 120
 Db 61 SDARKSETTITQSGYNGADVQDLVTRVVTHEMAHAFRNNATIDQWNAKNSDITVGVYGG 120
 QY 121 HEMAHANQTASDSSVMVRQVGFNNATANQY 151
 Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
 RESULT 13
 ID AAB36348
 AC AAB36348 standard; protein; 151 AA.
 XX
 XX AAB36348;
 DT 26-FEB-2001 (first entry)
 DE Agfa::PT3#3 amino acid sequence SEQ ID NO:16.
 KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX WO2000060102-A2.
 XX
 XX 12-OCT-2000.
 XX
 XX 05-APR-2000; 2000WO-CA000356.
 XX
 XX 05-APR-1999; 99US-0127888P.
 XX
 XX (UYVI-) UNIV VICTORIA.
 XX
 XX White AP, Doran JL, Collison SK, Kay WW;
 XX WPI; 2000-672631/65.
 XX

DR N-PSDB; AAC64624.
 XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 XX Disclosure; Page 136; 139pp; English.
 CC The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended are:
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 XX Sequence 151 AA;
 Query Match 74.2%; Score 575; DB 3; Length 151;
 Best Local Similarity 81.5%; Pred. No. 9.5e-49;
 Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
 Db 1 MKLLKVAFAAIVVSGSALAGVYDQLVTRVVTHEMAHAGSPDSTLSIYQYGSANAALALQ 60
 QY 61 SDARKSETTITQSGYNGADVGGADNSTIETQNGFRNNATIDQWNAKNDITVGVYGG 120
 Db 61 SDARKSETTITQSGYNGADVGGADNSTIETQNGFRNNATIDQWNAKNSDITVGVYGG 120
 QY 121 HEMAHANQTASDSSVMVRQVGFNNATANQY 151
 Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
 RESULT 14
 ID AAB36343
 AC AAB36343 standard; protein; 151 AA.
 XX
 XX AAB36343;
 DT 26-FEB-2001 (first entry)
 DE Escherichia coli CsgA amino acid sequence SEQ ID NO:7.
 KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 OS Escherichia coli.
 OS WO2000060102-A2.
 XX
 XX 12-OCT-2000.
 XX

PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
XX WPI; 2000-672631/65.
DR N-PSDB; AAC64619.
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 135; 139pp; English.
XX
XX The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SFP17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
XX the exemplification of the present invention
XX
SQ Sequence 151 AA;

Query Match 68.8%; Score 533; DB 3; Length 151;
Best Local Similarity 70.2%; Pred. No. 1.3e-44;
Matches 106; Conservative 18; Mismatches 27; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGGNGSGDPSTLSIYQYGSANAALQ 60
DB 1 MKLLKVAATAAIVFSGSALAGVVPQYGGGNGGNGSGPNSLNIVYQGGNSALQ 60

QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNATIDQWNAKNDQLVTRVVT 120
DB 61 TDARNSDLTITQGGNGADVGQSDSSIDLITQRFNGSATLDQWNGKNSMTVKQFGG 120

QY 121 HEMAHANQATSDSSVMVRQVGFNNATANQY 151
DB 121 GNGAAVDQTASNSVNVTVQVGFNNATAHQY 151

RESULT 15
ID ABR82651
XX ABR82651 standard; protein; 151 AA.
AC ABR82651;
XX
XX 04-DEC-2003 (first entry)
XX
XX E. coli CsgA subunit 15 kDa protein.
XX

KW Plasma protein; immune response; antibacterial; vaccine; gene therapy.
XX Escherichia coli.
XX
XX WO2003064446-A2.
XX
XX 07-AUG-2003.
XX
XX 30-JAN-2003; 2003WO-EP000943.
XX
XX 31-JAN-2002; 2002GB-00002275.
XX
XX (HANS-) HANSA MEDICAL RES AB.
XX
XX Bjoerck L, Olsen A, Wikstroem M, Herwald H;
XX
XX WPI; 2003-646136/61.
DR N-PSDB; ACF36153.
XX
XX New isolated peptide capable of binding a mammalian plasma protein,
PT useful in the manufacture of a medicament for the prevention and/or
PT treatment of a bacterial infection, such as Escherichia coli, Salmonella
PT or Shigella infections.
XX
XX Disclosure; Page 41-42; 42pp; English.
XX
XX The invention relates to an isolated peptide capable of binding a
CC mammalian plasma protein or of generating an immune response in a mammal
CC selected from sequences shown in ABR82642, ABR82648-49. The peptide or
CC antibody is useful for treating a bacterial infection in a human or
CC animal or in the manufacture of a medicament for the prophylactic
CC treatment of a bacterial infection, such as Escherichia coli, Salmonella
CC or Shigella infection. The peptide that is immobilized on a solid support
CC is also useful as a reagent for determining the ability of a plasma
CC protein to bind to bacteria. The present sequence represents an E. coli
CC 15 kDa protein
XX
SQ Sequence 151 AA;

Query Match 68.1%; Score 528; DB 7; Length 151;
Best Local Similarity 69.5%; Pred. No. 4.2e-44;
Matches 105; Conservative 18; Mismatches 28; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGGNGSGDPSTLSIYQYGSANAALQ 60
DB 1 MKLLKVEAIAAIVFSGSALAGVVPQYGGGNGGNGSGPNSLNIVYQGGNSALQ 60

QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNATIDQWNAKNDQLVTRVVT 120
DB 61 TDARNSDLTITQGGNGADVGQSDSSIDLITQRFNGSATLDQWNGKNSMTVKQFGG 120

QY 121 HEMAHANQATSDSSVMVRQVGFNNATANQY 151
DB 121 GNGAAVDQTASNSVNVTVQVGFNNATAHQY 151

Search completed: August 2, 2004, 14:48:27
Job time : 44.9 secs

This Page Blank (uspto)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:40:48 ; Search time 12 Seconds
(without alignments)
649.627 Million cell updates/sec

Title: US-09-543-407-24
Perfect score: 775
Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVGFNNATANY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/2/iaa/5A_COMP.pap.*
2: /cgn2_6/ptodata/2/iaa/5B_COMP.pap.*
3: /cgn2_6/ptodata/2/iaa/6A_COMP.pap.*
4: /cgn2_6/ptodata/2/iaa/6B_COMP.pap.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMP.pap.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	689	88.9	151	1	US-08-233-788A-59
2	505	65.2	120	1	US-08-233-788A-57
3	90.5	11.7	673	3	US-09-196-387-8
4	90.5	11.7	673	4	US-09-841-835-8
5	90.5	11.7	949	3	US-09-196-387-10
6	90.5	11.7	949	4	US-09-841-835-10
7	90.5	11.7	1327	3	US-09-196-387-2
8	90.5	11.7	1327	4	US-09-841-835-2
9	90.5	11.7	1327	3	US-09-972-115A-8
10	89.5	11.5	738	3	US-08-864-038A-3
11	87	11.2	1156	4	US-09-198-452A-171
12	84.5	10.9	339	4	US-09-252-991A-32096
13	83	10.7	889	4	US-09-336-447A-15
14	81.5	10.5	720	1	US-07-731-157A-2
15	81.5	10.5	720	2	US-08-541-780-2
16	80.5	10.4	389	4	US-09-134-001C-4316
17	80	10.3	2736	4	US-09-252-991A-30227
18	79.5	10.3	881	4	US-09-489-039A-12003
19	79	10.2	1415	4	US-09-252-991A-26438
20	79	10.2	2123	3	US-08-968-685A-10
21	78.5	10.1	589	4	US-09-489-039A-7849
22	78.5	10.1	745	4	US-09-336-115C-6
23	77	9.9	415	4	US-09-025-759B-280
24	77	9.9	444	1	US-07-881-075-3
25	77	9.9	444	1	US-08-120-827-3
26	77	9.9	444	1	US-08-478-675-3
27	77	9.9	1690	4	US-09-595-684B-39

28	76.5	9.9	227	4	US-09-328-352-7735	Sequence 7735, Ap
29	76	9.8	2680	4	US-08-489-039A-7973	Sequence 7973, Ap
30	75.5	9.7	238	4	US-09-495-880A-42	Sequence 42, Appl
31	75.5	9.7	645	4	US-09-919-172-41	Sequence 41, Appl
32	75.5	9.7	789	3	US-08-960-780-6	Sequence 6, Appl
33	75.5	9.7	789	3	US-09-073-898-6	Sequence 6, Appl
34	75.5	9.7	789	4	US-09-850-351A-6	Sequence 6, Appl
35	75.5	9.7	1036	4	US-09-206-942-73	Sequence 73, Appl
36	75.5	9.7	1051	4	US-09-134-001C-5005	Sequence 5005, Ap
37	75.5	9.7	1477	1	US-08-038-682-4	Sequence 4, Appl
38	75.5	9.7	1477	1	US-08-302-832-4	Sequence 4, Appl
39	75.5	9.7	1477	2	US-08-530-198-4	Sequence 4, Appl
40	75.5	9.7	1477	2	US-08-469-880-4	Sequence 4, Appl
41	75.5	9.7	1477	2	US-08-728-470-4	Sequence 4, Appl
42	75.5	9.7	1477	2	US-08-617-697-4	Sequence 4, Appl
43	75.5	9.7	1477	3	US-08-719-641-4	Sequence 4, Appl
44	75.5	9.7	1477	4	US-09-206-942-71	Sequence 71, Appl
45	75	9.7	211	1	US-08-276-852-34	Sequence 34, Appl

ALIGNMENTS

RESULT 1
US-08-233-788A-59
; Sequence 59, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, Karen S.
; APPLICANT: Clouthier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; TITLE OF INVENTION: OF SALMONELLA
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233,788A
FILING DATE: 26-APR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: King, Joshua
REGISTRATION NUMBER: 35,570
REFERENCE/DOCKET NUMBER: 920043.403C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEEDANBERRY
INFORMATION FOR SEQ ID NO: 59:
SEQUENCE CHARACTERISTICS:
LENGTH: 151 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-233-788A-59

Query Match 88.9%; Score 689; DB 1; Length 151;
Best Local Similarity 91.4%; Pred. No. 1.8e-60;
Matches 139; Conservative 2; Mismatches 11; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNNHNGNSSGPDSTLSIIYQYGSANAALQ 60
|||||

Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALQ 60
 QY 61 SPARKSETTITQSGYNGADVGCGADNSTIETQNGFRNNATIDQWNAKNDQLVTRVVT 120
 Db 61 SPARKSETTITQSGYNGADVGCGADNSTIETQNGFRNNATIDQWNAKNSDITVQYGG 120
 QY 121 HEMAHANQASDSSVMVRQVGGNNATQY 151
 Db 121 NNPALVNTQASDSSVMVRQVGGNNATQY 151

RESULT 2

US-08-233-788A-57
 ; Sequence 57, Application US/08233788A
 ; Patent No. 5635617
 ; GENERAL INFORMATION:
 ; APPLICANT: Doran, James L.
 ; APPLICANT: Kay, William W.
 ; APPLICANT: Collinson, Karen S.
 ; APPLICANT: Cloutier, Sharon C.
 ; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
 ; TITLE OF INVENTION: OF SALMONELLA
 ; NUMBER OF SEQUENCES: 61
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Seed and Berry
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue
 ; CITY: Seattle
 ; STATE: Washington
 ; COUNTRY: U.S.A.
 ; ZIP: 98104-7092
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/233,788A
 ; FILING DATE: 26-APR-1994
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: King, Joshua
 ; REGISTRATION NUMBER: 35,570
 ; REFERENCE/DOCKET NUMBER: 920043.403C2
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (206) 622-4900
 ; TELEFAX: (206) 682-6031
 ; TELEX: 3723836 SEEDANBERRY
 ; INFORMATION FOR SEQ ID NO: 57:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 120 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-233-788A-57

Query Match 65.2%; Score 505; DB 1; Length 120;
 Best Local Similarity 88.4%; Pred. No. 1.7e-42;
 Matches 99; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

QY 22 VVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALQSDARKSETTITQSGYNGADV 81
 Db 1 VVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALQSDARKSETTITQSGYNGADV 60
 QY 82 GQADNSTIETQNGFRNNATIDQWNAKNDQLVTRVTHEMAHANQASDS 133
 Db 61 GQADNSTIETQNGFRNNATIDQWNAKNSDITVQYGGNNAALVNTQASDS 112

RESULT 3

US-09-196-387-8
 ; Sequence 8, Application US/09196387
 ; Patent No. 6277613
 ; GENERAL INFORMATION:

; APPLICANT: de Lange, Titia
 ; APPLICANT: Smith, Susan
 ; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
 ; TITLE OF INVENTION: OF USE THEREOF
 ; NUMBER OF SEQUENCES: 12
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Klauber & Jackson
 ; STREET: 411 Hackensack Avenue, 4th Floor
 ; CITY: Hackensack
 ; STATE: New Jersey
 ; COUNTRY: USA
 ; ZIP: 07601
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/196,387
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 09/095,225
 ; FILING DATE: June 10, 1998
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Jackson Esq., David A.
 ; REGISTRATION NUMBER: 26,742
 ; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 201-487-5800
 ; TELEFAX: 201-343-1684
 ; TELEX: 133521
 ; INFORMATION FOR SEQ ID NO: 8:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 673 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-09-196-387-8

Query Match 11.7%; Score 90.5; DB 3; Length 673;
 Best Local Similarity 28.4%; Pred. No. 1.1;
 Matches 42; Conservative 16; Mismatches 61; Indels 29; Gaps 6;

QY 6 VAAFAAI-VVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALQSDAR 64
 Db 99 VAAAPVVPVAVTSSAAGVAPNPAGSGNNKNSPSSSSPTSS-SSSSPSPGSSLAESPEAA 157
 QY 65 KSETTIT-----QSGYNGADVGGADNSTIETQNGFRNNATIDQWNAKNDQLVTRVVT 120
 Db 158 GVSSTAPLPGGAAGPGTGVPAVSGALRELEACRG-----DVSRRVKRLVDA--- 204
 QY 121 HEMAHANQASD-----SSVMVRQVGF 143
 Db 205 -----ANVNAKDVAGRKSSPLHFAAGFG 227

RESULT 4
 US-09-841-835-8
 ; Sequence 8, Application US/09841835
 ; Patent No. 6506587
 ; GENERAL INFORMATION:
 ; APPLICANT: de Lange, Titia
 ; APPLICANT: Smith, Susan
 ; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
 ; TITLE OF INVENTION: OF USE THEREOF
 ; NUMBER OF SEQUENCES: 12
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Klauber & Jackson
 ; STREET: 411 Hackensack Avenue, 4th Floor
 ; CITY: Hackensack
 ; STATE: New Jersey
 ; COUNTRY: USA


```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 949 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-841-835-10

Query Match 11.7%; Score 90.5; DB 4; Length 949;
Best Local Similarity 28.4%; Pred. No. 1.8;
Matches 42; Conservative 16; Mismatches 61; Indels 29; Gaps 6;

QY 6 VAAPVPAVSTSSAAGVAPNPAAGSGNNHNGGNSGPDSTLSIYQYGSANAALQSDAR 64
Db 99 VAAAPVPAVSTSSAAGVAPNPAAGSGNNHNGGNSGPDSTLSIYQYGSANAALQSDAR 157

QY 65 KSETTIT---QSGYNGADVGQADNSTIETQNGFRNNATIDQWNAKNDQLVTRVVT 120
Db 158 GVSSTAPLPGGAAGPTGTPAVSGALRELLACRNG-----DVSrvKLVDA--- 204

QY 121 HEMAHANQTASD-----SSVMVROVQVGF 143
Db 205 -----ANVNAKDMAGRKSSPLHFAAGFG 227

RESULT 7
US-09-196-387-2
; Sequence 2, Application US/09196387
; Patent No. 6277613
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; TITLE OF INVENTION: OF USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/196,387
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/095,225
; FILING DATE: June 10, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1327 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
US-09-196-387-2

Query Match 11.7%; Score 90.5; DB 3; Length 1327;
Best Local Similarity 28.4%; Pred. No. 2.8;
Matches 42; Conservative 16; Mismatches 61; Indels 29; Gaps 6;

QY 6 VAAPVPAVSTSSAAGVAPNPAAGSGNNHNGGNSGPDSTLSIYQYGSANAALQSDAR 64
Db 99 VAAAPVPAVSTSSAAGVAPNPAAGSGNNHNGGNSGPDSTLSIYQYGSANAALQSDAR 157

QY 65 KSETTIT---QSGYNGADVGQADNSTIETQNGFRNNATIDQWNAKNDQLVTRVVT 120
Db 158 GVSSTAPLPGGAAGPTGTPAVSGALRELLACRNG-----DVSrvKLVDA--- 204

US-09-841-835-2
; Sequence 2, Application US/09841835
; Patent No. 6506587
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; TITLE OF INVENTION: OF USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/841,835
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/196,387
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1327 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
US-09-841-835-2

Query Match 11.7%; Score 90.5; DB 4; Length 1327;
Best Local Similarity 28.4%; Pred. No. 2.8;
Matches 42; Conservative 16; Mismatches 61; Indels 29; Gaps 6;

QY 6 VAAPVPAVSTSSAAGVAPNPAAGSGNNHNGGNSGPDSTLSIYQYGSANAALQSDAR 64
Db 99 VAAAPVPAVSTSSAAGVAPNPAAGSGNNHNGGNSGPDSTLSIYQYGSANAALQSDAR 157

QY 65 KSETTIT---QSGYNGADVGQADNSTIETQNGFRNNATIDQWNAKNDQLVTRVVT 120
Db 158 GVSSTAPLPGGAAGPTGTPAVSGALRELLACRNG-----DVSrvKLVDA--- 204
```


QY 121 HEMAHANOTASD-----SSVMVRQVGF 143
Db 205 -----ANVNADMAGKSSPLHFAAGF 227

RESULT 9

US-09-972-115A-8
; Sequence 9, Application US/09972115A
; Patent No. 6599728
; GENERAL INFORMATION:
; APPLICANT: Geron Corporation
; APPLICANT: Gregg, Morin B.
; APPLICANT: Walter, Funk D.
; APPLICANT: Mieczyslaw, Flatuszek A.
; FILE REFERENCE: 080/003C
; CURRENT APPLICATION NUMBER: US/09/972,115A
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: US 60/128,577
; PRIOR FILING DATE: 2000-04-10
; PRIOR APPLICATION NUMBER: US 60/129,123
; PRIOR FILING DATE: 1999-04-13
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 1327
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-972-115A-8

Query Match 11.7%; Score 90.5; DB 4; Length 1327;
Best Local Similarity 28.4%; Pred. No. 2.8;
Matches 42; Conservative 16; Mismatches 61; Indels 29; Gaps 6;
QY 6 VAFAAI-VVGSALAGVVPWGGGNGHNGSGDPSTLSIYQGSANAALALQSDAR 64
Db 99 VAAAPVPAVSTSSAAGVAPNAGSGNNSPSSSSPTSS-SSSSPSGSSLAESPEAA 157
QY 65 KSETTIT---QSGYNGADVQCGADNSTILTQNGFNNATIDQWNAKNDQLATRVVT 120
Db 158 GVSSTAPIGPAGAGPGTGPVAVSGALRELLACRNG-----DVSrvKRLVDA--- 204
QY 121 HEMAHANOTASD-----SSVMVRQVGF 143
Db 205 -----ANVNADMAGKSSPLHFAAGF 227

RESULT 10

US-08-864-038A-3
; Sequence 3, Application US/08864038A
; Patent No. 6001592
; GENERAL INFORMATION:
; APPLICANT: Kunio NAKASHIMA et al.
; TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE CDNA, VECTOR
; TITLE OF INVENTION: CONTAINING SAID CDNA, HOST CELLS TRANSFORMED WITH SAID
; TITLE OF INVENTION: VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING
; TITLE OF INVENTION: SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBODY
; TITLE OF INVENTION: TO SAID POLYPEPTIDE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESS: 812-5 Hirano
; STREET: Isshinden
; CITY: Tsu-city
; STATE: Mie-prefecture
; COUNTRY: JAPAN
; ZIP: 514-01
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Microsoft Windows 95
; SOFTWARE: Word Perfect 6.1
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/864,038A
; FILING DATE: May 28, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-184459
; FILING DATE: 15-July-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: C. Bruce Hamburg
; REGISTRATION NUMBER: 22,389
; REFERENCE/DOCKET NUMBER: F-5610
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)986-2340
; TELEFAX: (212)953-7733
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 738
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Pinctada fucata
; CELL TYPE: mantle epithelial cell
; FEATURE:
; NAME/KEY: peptide
; LOCATION: from 1 to 738
; IDENTIFICATION METHOD: E (by experiment)
US-08-864-038A-3

Query Match 11.5%; Score 89.5; DB 3; Length 738;
Best Local Similarity 35.4%; Pred. No. 1.6;
Matches 29; Conservative 4; Mismatches 26; Indels 23; Gaps 3;
QY 3 LLKVAFAAIVVGSALAGVVPWGGGNGHNGSGSGDPSTLSIYQGSANAALALQSD 62
Db 419 LLKSSASASASASASAG-----GGGGGNGGGGGG-----GGGAGALA----- 460
QY 63 ARKSETTITQSGYNGADVQCG 84
Db 461 -----AALAAAGAGGGLGGGG 477

RESULT 11

US-09-198-452A-171
; Sequence 171, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Griffais, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
; TITLE OF INVENTION: and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 171
; LENGTH: 1156
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-09-198-452A-171

Query Match 11.2%; Score 87; DB 4; Length 1156;
Best Local Similarity 32.0%; Pred. No. 5.1;
Matches 32; Conservative 14; Mismatches 34; Indels 20; Gaps 5;
QY 44 TLSIYQGSANAALALQSDARKSETTITQSGYNGADVQCGADNSTI-----ELTQ 94
Db 859 TVSPEDYAAVQAALA--AVVRKHESLIV-STYGLGAQEGQTSSKVTTLMDLHAEELVE 915
QY 95 NGERNNATIDQWNAKNDQLATRVVTHEMAHANQTASDSS 134
Db 916 MG-----VETRLNRSDQILHRV--HSVLHSHLRDSDSS 947

RESULT 12

```

US-09-252-991A-32096
; Sequence 32096, Application US/09252591A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 32096
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-32096

Query Match 10.9%; Score 84.5; DB 4; Length 339;
Best Local Similarity 21.5%; Pred. No. 1.8;
Matches 34; Conservative 32; Mismatches 67; Indels 25; Gaps 5;

QY 5 KVAFAAIVVSGSALAGVVPQWGG-----GGNHGGGSSGPDSTLSIYQYGSANAAL 57
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 58 KVSFNGTLL--NNASVSGSIKIDAGNVGNVAAAGDNNQQAALASADASFVGTATAS- 114
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 58 ALQSDARKSETTITQSGYNGGADVGGADNSTIELTQNGFRNATIDQWNAKYNDO---- 113
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 115 -----TSVLOGYGTNTLVNPNSTLSNSANNVSGNLGVNV-AGNFNQKND 163
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 114 LVTRVVTHEMAHANTASDVSVVRQVGFNNATANQY 151
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 164 LAAAVSNGQYSTAGSAASQTSIGNTVNSANVAYGTY 201
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 13
US-09-336-447A-15
; Sequence 15, Application US/09336447A
; Patent No. 6310190
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USP41 AND USP42 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY.024
; CURRENT APPLICATION NUMBER: US/09/336,447A
; CURRENT FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 15
; LENGTH: 889
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-336-447A-15

Query Match 10.7%; Score 83; DB 4; Length 889;
Best Local Similarity 27.9%; Pred. No. 8.9;
Matches 39; Conservative 16; Mismatches 51; Indels 34; Gaps 8;

QY 28 GGNHGGGSSGPDSTLSIYQ---YGSANALALQSDARKSETTITQSGY-----GNGA 79
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 69 GAGRHNNGGSAHHSILGGWNTVNGYTSALVGGYGNETQDGYTFVGGGVNKLAKGNYT 128
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 80 DVGCG-----ADNSTIELTQNGFRN-----NATID---QWNAKYNDOQLVTRVVTHEMAH 125
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 129 FVGGYKNLAEGDNATI---AGGFANLASEGDNATAGGFENRAGSIVSWSG-----GY 179
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

QY 126 ANQTASDSSVVRQVGFNN 145
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 180 ANQATGESSTV---AGGSNN 196
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 14
US-07-731-157A-2
; Sequence 2, Application US/07731157A
; Patent No. 5457032
; GENERAL INFORMATION:
; APPLICANT: Quax, Wilhelmus J.
; APPLICANT: Misset, Onno
; APPLICANT: Van der Laan, Jan M.
; APPLICANT: Lenthing, Herman B.M.
; TITLE OF INVENTION: Mutated beta-lactam acylase genes
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: COOLEY GODWARD CASTRO HUDDLESON & TATUM
; STREET: FIVE PALO ALTO SQUARE, 4TH FLOOR
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/731,157A
; FILING DATE: 19910509
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 90200962
; FILING DATE: 18-APR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: RAE-VENTER PH.D., BARBARA
; REGISTRATION NUMBER: 32,750
; REFERENCE/DOCKET NUMBER: GBRO-027/00US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-494-7622
; TELEFAX: 415-857-0663
; TELEX: 380816 COOLEY PA
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 720 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-731-157A-2

Query Match 10.5%; Score 81.5; DB 1; Length 720;
Best Local Similarity 25.8%; Pred. No. 9.5;
Matches 46; Conservative 17; Mismatches 56; Indels 59; Gaps 9;

QY 1 MKLLKVAFAAIVVSGSALAGVV-----PQ-----WGGGNGHGGG 36
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 2 LRVLHRAASALVMATVIGLAPAVAFALABPTSTPQAPIAAYKPRSNELWDG---YGV 57
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 37 NSSGPDSTLSIYQYGSANA-----ALALQSDARKSETTITQSGYNGGADV-GQGADNST 89
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 58 HIYGVDAPSAFYGYGWAQARSQGDNLRLYGEAR-----GKGEYWGPDYEQT 106
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 90 IELTQNGFRNATIDQWNAK-----NYDQLVTRVVTHEMAHANTASDSSVVRQV 140
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 107 VLLLTNGVPERA--QQWYAAQQSPDFRANLDAFAAGI---NAYAAQQNPDISPDVRQV 158
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 15
US-08-541-780-2
; Sequence 2, Application US/08541780
; Patent No. 5935831
; GENERAL INFORMATION:

```

APPLICANT: Quax, Wilhelmus J.
APPLICANT: Misset, Onno
APPLICANT: Van der Laan, Jan M.
TITLE OF INVENTION: Mutated beta-lactam acylase genes
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: COOLEY GODWARD CASTRO HUDDLESON & TATUM
STREET: FIVE PALO ALTO SQUARE, 4TH FLOOR
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPC)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/541,780
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/731,157
FILING DATE:
APPLICATION NUMBER: EP 90200962
FILING DATE: 18-APR-1990
ATTORNEY/AGENT INFORMATION:
NAME: RAE-VENTER PH.D., BARBARA
REGISTRATION NUMBER: 32,750
REFERENCE/DOCKET NUMBER: GERO-027/000S
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-494-7622
TELEFAX: 415-857-0663
TELEX: 380816 COOLEY PA
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 720 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-541-780-2

Query Match 10.5%; Score 81.5; DB 2; Length 720;
Best Local Similarity 25.8%; Pred. No. 9.5;
Matches 46; Conservative 17; Mismatches 56; Indels 59; Gaps 9;
QY 1 MKLLKVAFAAIVVSGSALAGV-----PQ-----WCGGNGHNGG 36
Db 2 LRVLRASALVMTVIGLAPAVAFALAEPTSTQAPIAAYKPRSNEILMDG----YGV 57
QY 37 NSSGPDSTLSIYQGSANA-----ALALQSDARKSETTITQSGYNGADV-GQADNST 89
Db 58 HIYGVDAFSAFYGYGWAQARSQGDNIILRYGEAR-----GKGAEWGPDYEQT 106
QY 90 IELTONGERNATIDWNAK-----NYDOLVTRVVTNVAHANQTASSSVMVROV 140
Db 107 VLLLTNGVPERA--QQWYAQQSPDFRNLDAFAAGI-----NAYAQQNPDDISPDVRQV 158

Search completed: August 2, 2004, 14:58:35
Job time : 12 secs

This Page Blank (uspto)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:54:48 ; Search time 36.8 Seconds

(without alignments)
1287.123 Million cell updates/sec

Title: US-09-543-407-24

Perfect score: 775

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVGFNNATANYQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	530	68.4	151	12	US-09-741-873B-4
2	530	68.4	151	12	US-09-741-873B-4
3	452	58.3	131	12	US-09-741-873B-2
4	452	58.3	131	12	US-09-741-873B-2
5	111	14.3	445	15	US-10-369-493-20638
6	91	11.7	688	14	US-10-032-585-7876
7	90.5	11.7	263	12	US-10-425-114-49960
8	90.5	11.7	673	9	US-09-841-835-10
9	90.5	11.7	949	9	US-09-841-835-2
10	90.5	11.7	1327	9	US-09-841-835-2
11	90.5	11.7	1327	10	US-09-972-115A-8
12	90.5	11.7	1327	14	US-10-159-937-4
13	89	11.5	369	12	US-10-425-114-56041
14	89	11.5	486	12	US-10-424-599-275468
15	89	11.5	507	12	US-10-425-114-57763

16	88.5	11.4	145	16	US-10-437-963-147748	Sequence 147748,
17	88.5	11.4	438	14	US-10-156-761-9343	Sequence 9343, Ap
18	87	11.2	1156	15	US-10-289-762-171	Sequence 171, App
19	86	11.1	186	12	US-10-282-122A-49412	Sequence 49412, A
20	86	11.1	482	14	US-10-156-761-8763	Sequence 8763, Ap
21	86	11.1	762	12	US-10-282-122A-59874	Sequence 59874, A
22	85.5	11.0	270	16	US-10-437-963-122263	Sequence 122263,
23	85.5	11.0	271	14	US-10-156-761-11721	Sequence 11721, A
24	85.5	11.0	2457	12	US-10-282-122A-49854	Sequence 49854, A
25	85	11.0	478	16	US-10-437-963-115033	Sequence 115033,
26	84	10.8	254	10	US-09-880-748-1136	Sequence 1136, Ap
27	84	10.8	254	10	US-09-880-748-1136	Sequence 1136, Ap
28	84	10.8	254	12	US-10-293-418-1136	Sequence 1136, Ap
29	84	10.8	254	12	US-10-293-418-1136	Sequence 1136, Ap
30	83.5	10.8	350	14	US-10-156-761-13168	Sequence 13168, A
31	83.5	10.8	394	16	US-10-437-963-163379	Sequence 163379,
32	83	10.7	242	12	US-10-425-114-61520	Sequence 61520, A
33	83	10.7	889	10	US-09-952-267-15	Sequence 15, Appl
34	82.5	10.6	1074	12	US-10-282-122A-50616	Sequence 50616, A
35	82	10.6	214	16	US-10-437-963-119143	Sequence 119143,
36	82	10.6	254	10	US-09-880-748-1226	Sequence 1226, Ap
37	82	10.6	254	12	US-10-293-418-1226	Sequence 1226, Ap
38	82	10.6	678	12	US-10-282-122A-64573	Sequence 64573, A
39	81.5	10.5	883	12	US-10-282-122A-42626	Sequence 42626, A
40	81.5	10.5	189	16	US-10-437-963-170736	Sequence 170736,
41	81.5	10.5	256	14	US-10-156-761-11972	Sequence 11972, A
42	81.5	10.5	429	16	US-10-437-963-124678	Sequence 124678,
43	81.5	10.5	449	12	US-10-424-599-285485	Sequence 285485,
44	81.5	10.5	509	12	US-10-425-114-49393	Sequence 48393, A
45	81.5	10.5	529	12	US-10-425-114-49406	Sequence 49406, A

ALIGNMENTS

RESULT 1

US-09-741-873B-4
; Sequence 4, Application US/09741873B
; Publication No. US20020081722A1
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olesen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; CURRENT FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-4

Query Match 68.4%; Score 530; DB 12; Length 151;

Best Local Similarity 69.8%; Pred. No. 7.2e-46;

Matches 105; Conservative 19; Mismatches 27; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNHNGGNSGPDSTLSIYQYGSANAALQ 60

[illegible]

```

RESULT 2
US-09-741-873B-4
; Sequence 4, Application US/09741873B
; Publication No. US20040096965A9
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741.873B

```

[illegible]

```

RESULT 3
US-09-741-873B-2
; Sequence 2, Application US/09741873B
; Publication NO. US20020081722A1
; GENERAL INFORMATION:
; APPLICANT: Nozmark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; CURRENT FILING DATE: 2003-04-04

```

```

, PRIOR APPLICATION NUMBER: SE 8801723-1
, PRIOR FILING DATE: 1998-05-06
, PRIOR APPLICATION NUMBER: US 08/978,878
, PRIOR FILING DATE: 1997-11-26
, PRIOR APPLICATION NUMBER: US 07/347,189
, PRIOR FILING DATE: 1989-05-04
, PRIOR APPLICATION NUMBER: US 07/789,437
, PRIOR FILING DATE: 1991-11-06
, PRIOR APPLICATION NUMBER: US 07/970,846
, PRIOR FILING DATE: 1992-11-03
, PRIOR APPLICATION NUMBER: US 08/187,865
, PRIOR FILING DATE: 1994-01-28
, PRIOR APPLICATION NUMBER: US 08/318,519
, PRIOR FILING DATE: 1994-10-05
, NUMBER OF SEQ ID NOS: 10
, SOFTWARE: PatentIn version 3.0
, SEQ ID NO 2
, LENGTH: 131
, TYPE: PRT
, ORGANISM: Escherichia coli
, US-09-741-873B-2

```

Query Match	58.3%;	Score	452;	DB	12;	Length	131;
Best Local Similarity	66.4%;	Pred. No.	5e-38;				
Matches	87;	Conservative	18;	Mismatches	26;	Indels	0;
Gaps	0						
Qy	21	GVVPQWGGGNGHGGNGSSGPDSTLSIVQVGSANAALQSDARKGETTITQSGYGGNGAD	80				
Db	1	GVVPQVQGGGNGHGGGNGNSGPNSELTYYQVGGGNSALALQTDARNSDLTITQHGCGGNGAD	50				
Qy	81	VGQGDNSTIELTQNGFFNNATIDQWAKNYDQLTVFWTTHMAHANQATSSVNVVQV	140				
Db	61	VGQGDSDSSIDLTQRFNSATLQWNGKQSEMTVKQFGGCGAAVDQTASNSVNVTVQV	120				
Qy	141	GFGNNTANQY	151				
Db	121	GFGNNTAHQY	131				

```

RESULT 4
US-09-741-873B-2
; Sequence 2, Application US/09741873B
; Publication NO. US20040096965A9
; GENERAL INFORMATION:
; APPLICANT: Norman, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; CURRENT FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-2
Query Match          58.3%; Score 452; DB 12; Length 131;

```



```
; TELEPAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1327 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
US-09-841-835-2

Query Match      11.7%; Score 90.5; DB 9; Length 1327;
Best Local Similarity 28.4%; Pred. No. 4.8;
Matches 42; Conservative 16; Mismatches 61; Indels 29; Gaps 6;

QY 6 VAAFAAI-VVSGSALAGVVPQWGGGNGGNSGPDSTLSIYQYGSANAALALQSDAR 64
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 99 VAAAPVPAVSTSSAAGVAPNPAGSGNNSPSSSTSS-SSSSPSPGSSLAESPEAA 157
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 65 KSETTIT- - - - -QSGYGNAGDVGGADNSTIETQNGFRNNATIDQWNAKNDQLVTRVVT 120
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 158 GVSSTAPLPGAAGTGTGVPVAVSGALRELLEACRNG- - - - -DVSrvKRLVDA- - - 204

QY 121 HEMAHANOTASD- - - - -SSVMVRQVGF 143
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 205 - - - - -ANVNAKDMAGRKSSPLHFAAGFG 227

RESULT 11
US-09-972-115A-8
; Sequence 8, Application US/09972115A
; Publication No. US20030032769A1
; GENERAL INFORMATION:
; APPLICANT: Geron Corporation
; APPLICANT: Gregg, Morin B.
; APPLICANT: Walter, Funk D.
; APPLICANT: Mieczyslaw, Piatyszek A.
; TITLE OF INVENTION: A Second Mammalian Telomerase
; FILE REFERENCE: 080/003C
; CURRENT APPLICATION NUMBER: US/09/972,115A
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: US 60/128,577
; PRIOR FILING DATE: 2000-04-10
; PRIOR APPLICATION NUMBER: US 60/129,123
; PRIOR FILING DATE: 1999-04-13
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 1327
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-972-115A-8

Query Match      11.7%; Score 90.5; DB 10; Length 1327;
Best Local Similarity 28.4%; Pred. No. 4.8;
Matches 42; Conservative 16; Mismatches 61; Indels 29; Gaps 6;

QY 6 VAAFAAI-VVSGSALAGVVPQWGGGNGGNSGPDSTLSIYQYGSANAALALQSDAR 64
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 99 VAAAPVPAVSTSSAAGVAPNPAGSGNNSPSSSTSS-SSSSPSPGSSLAESPEAA 157
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 65 KSETTIT- - - - -QSGYGNAGDVGGADNSTIETQNGFRNNATIDQWNAKNDQLVTRVVT 120
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 158 GVSSTAPLPGAAGTGTGVPVAVSGALRELLEACRNG- - - - -DVSrvKRLVDA- - - 204

QY 121 HEMAHANOTASD- - - - -SSVMVRQVGF 143
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 205 - - - - -ANVNAKDMAGRKSSPLHFAAGFG 227

RESULT 12
US-10-199-937-4
; Sequence 4, Application US/10199937
; Publication No. US20030190739A1
; GENERAL INFORMATION:
; APPLICANT: Christenson, Erik
; APPLICANT: Demaggio, Anthony J.
; APPLICANT: Goldman, Phyllis S.
; APPLICANT: McElligott, David L.
; TITLE OF INVENTION: TANKYRASE2 MATERIALS AND METHODS
; FILE REFERENCE: 27866/36559
; CURRENT APPLICATION NUMBER: US/10/199,937
; CURRENT FILING DATE: 2002-07-22
; PRIOR APPLICATION NUMBER: US/09/606,035
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 60/141,582
; PRIOR FILING DATE: 1998-06-29
; NUMBER OF SEQ ID NOS: 178
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1327
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-199-937-4

Query Match      11.7%; Score 90.5; DB 14; Length 1327;
Best Local Similarity 28.4%; Pred. No. 4.8;
Matches 42; Conservative 16; Mismatches 61; Indels 29; Gaps 6;

QY 6 VAAFAAI-VVSGSALAGVVPQWGGGNGGNSGPDSTLSIYQYGSANAALALQSDAR 64
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 99 VAAAPVPAVSTSSAAGVAPNPAGSGNNSPSSSTSS-SSSSPSPGSSLAESPEAA 157
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 65 KSETTIT- - - - -QSGYGNAGDVGGADNSTIETQNGFRNNATIDQWNAKNDQLVTRVVT 120
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 158 GVSSTAPLPGAAGTGTGVPVAVSGALRELLEACRNG- - - - -DVSrvKRLVDA- - - 204

QY 121 HEMAHANOTASD- - - - -SSVMVRQVGF 143
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 205 - - - - -ANVNAKDMAGRKSSPLHFAAGFG 227

RESULT 13
US-10-425-114-56041
; Sequence 56041, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 56041
; LENGTH: 369
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: 701205720_FLI.pep
US-10-425-114-56041

Query Match      11.5%; Score 89; DB 12; Length 369;
Best Local Similarity 26.3%; Pred. No. 1.4;
Matches 36; Conservative 20; Mismatches 61; Indels 20; Gaps 6;

QY 14 VSGSALAGVVPQWGGGNGGNSGPDSTLSIYQYGSANAALALQSDARKSETTITQS 73
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 13 VRSSASKG-KPPLSAGNNNGGWDNDPDD- - - - -GFGSARGG- - - - -ADLRNQSTGQDVR 62
```

QY 74 GYNGADVGGADNSTIETQNGFRNNATIDQWNAKXYDQLVTRVVVTHEMAHANQTASDS 133
Db 63 GFGGGNNV-----SRSKSTQDMY-TRALEASAANKEDFFARKRAENESRPEGLPPSQG 115

QY 134 SVMVROVGFQGNATANQ 150
Db 116 G---KYVGFSGGPAPNQ 129

RESULT 14

US-10-424-599-275468
; Sequence 275468, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 275468
; LENGTH: 486
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_90769C.1.pap
US-10-424-599-275468

Query Match 11.5%; Score 89; DB 12; Length 486;
Best Local Similarity 26.3%; Pred. No. 1.9;
Matches 36; Conservative 20; Mismatches 61; Indels 20; Gaps 6;

QY 14 VSGSALAGVVPQWGGGNNHNGGNSGSPDSTLSIYQYGSANAALALQSDARKSETTITQS 73
Db 130 VRSSASKG-KPPLSAGNNNGGNDWDPDD-----GFGSARGG-----ADLRNQSTGDVR 179

QY 74 GYNGADVGGADNSTIETQNGFRNNATIDQWNAKXYDQLVTRVVVTHEMAHANQTASDS 133
Db 180 GFGGGNNV-----SRSKSTQDMY-TRALEASAANKEDFFARKRAENESRPEGLPPSQG 232

QY 134 SVMVROVGFQGNATANQ 150
Db 233 G---KYVGFSGGPAPNQ 246

RESULT 15

US-10-425-114-57763
; Sequence 57763, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 57763
; LENGTH: 507
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-GMFLMINSOY049H04_FLI.pap
US-10-425-114-57763

Query Match 11.5%; Score 89; DB 12; Length 507;
Best Local Similarity 26.3%; Pred. No. 2;
Matches 36; Conservative 20; Mismatches 61; Indels 20; Gaps 6;
QY 14 VSGSALAGVVPQWGGGNNHNGGNSGSPDSTLSIYQYGSANAALALQSDARKSETTITQS 73
Db 151 VRSSASKG-KPPLSAGNNNGGNDWDPDD-----GFGSARGG-----ADLRNQSTGDVR 200
QY 74 GYNGADVGGADNSTIETQNGFRNNATIDQWNAKXYDQLVTRVVVTHEMAHANQTASDS 133
Db 201 GFGGGNNV-----SRSKSTQDMY-TRALEASAANKEDFFARKRAENESRPEGLPPSQG 253
QY 134 SVMVROVGFQGNATANQ 150
Db 254 G---KYVGFSGGPAPNQ 267

Search completed: August 2, 2004, 15:36:12
Job time : 36.8 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:48:33 ; Search time 167.9 Seconds
(without alignments)
877.809 Million cell updates/sec

Title: US-09-543-407-24
Perfect score: 775
Sequence: 1 MKLLKVAFAAIVVGSALA.....DSVVRQVGFNNATANQY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 6019581 seqs, 976053577 residues
Total number of hits satisfying chosen parameters: 6019581

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents AA.Main:*
1: /cgn2_6/ptodata/2/paa/PTCTUS_COMB.pcp.*
2: /cgn2_6/ptodata/2/paa/US06_COMB.pcp.*
3: /cgn2_6/ptodata/2/paa/US07_COMB.pcp.*
4: /cgn2_6/ptodata/2/paa/US080_COMB.pcp.*
5: /cgn2_6/ptodata/2/paa/US081_COMB.pcp.*
6: /cgn2_6/ptodata/2/paa/US082_COMB.pcp.*
7: /cgn2_6/ptodata/2/paa/US083_COMB.pcp.*
8: /cgn2_6/ptodata/2/paa/US084_COMB.pcp.*
9: /cgn2_6/ptodata/2/paa/US085_COMB.pcp.*
10: /cgn2_6/ptodata/2/paa/US086_COMB.pcp.*
11: /cgn2_6/ptodata/2/paa/US087_COMB.pcp.*
12: /cgn2_6/ptodata/2/paa/US088_COMB.pcp.*
13: /cgn2_6/ptodata/2/paa/US089_COMB.pcp.*
14: /cgn2_6/ptodata/2/paa/US090_COMB.pcp.*
15: /cgn2_6/ptodata/2/paa/US091_COMB.pcp.*
16: /cgn2_6/ptodata/2/paa/US092_COMB.pcp.*
17: /cgn2_6/ptodata/2/paa/US093_COMB.pcp.*
18: /cgn2_6/ptodata/2/paa/US094_COMB.pcp.*
19: /cgn2_6/ptodata/2/paa/US095_COMB.pcp.*
20: /cgn2_6/ptodata/2/paa/US096_COMB.pcp.*
21: /cgn2_6/ptodata/2/paa/US097A_COMB.pcp.*
22: /cgn2_6/ptodata/2/paa/US097B_COMB.pcp.*
23: /cgn2_6/ptodata/2/paa/US098_COMB.pcp.*
24: /cgn2_6/ptodata/2/paa/US099A_COMB.pcp.*
25: /cgn2_6/ptodata/2/paa/US099B_COMB.pcp.*
26: /cgn2_6/ptodata/2/paa/US100_COMB.pcp.*
27: /cgn2_6/ptodata/2/paa/US101_COMB.pcp.*
28: /cgn2_6/ptodata/2/paa/US102_COMB.pcp.*
29: /cgn2_6/ptodata/2/paa/US103_COMB.pcp.*
30: /cgn2_6/ptodata/2/paa/US104_COMB.pcp.*
31: /cgn2_6/ptodata/2/paa/US106_COMB.pcp.*
32: /cgn2_6/ptodata/2/paa/US107_COMB.pcp.*
33: /cgn2_6/ptodata/2/paa/US60_COMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
------------	-------	-------	--------	----	-------------

1	775	100.0	151	19	US-09-543-407-24	Sequence 24, Appl
2	714	92.1	151	19	US-09-543-407-14	Sequence 14, Appl
3	700	90.3	151	19	US-09-543-407-30	Sequence 30, Appl
4	690	89.0	151	19	US-09-543-407-5	Sequence 5, Appl
5	689	88.9	151	6	US-08-233-642A-57	Sequence 57, Appl
6	655	84.5	151	19	US-09-543-407-12	Sequence 12, Appl
7	609	78.6	151	19	US-09-543-407-22	Sequence 22, Appl
8	609	78.6	151	19	US-09-543-407-26	Sequence 26, Appl
9	608	78.5	151	19	US-09-543-407-18	Sequence 18, Appl
10	606	78.2	151	19	US-09-543-407-20	Sequence 20, Appl
11	603	77.8	131	19	US-09-543-407-31	Sequence 31, Appl
12	600	77.4	131	19	US-09-543-407-28	Sequence 28, Appl
13	575	74.2	151	19	US-09-543-407-16	Sequence 16, Appl
14	533	68.8	151	19	US-09-543-407-7	Sequence 7, Appl
15	530	68.4	151	13	US-08-978-878-4	Sequence 4, Appl
16	530	68.4	151	21	US-09-741-873B-4	Sequence 4, Appl
17	528	68.1	151	33	US-60-352-946-2	Sequence 2, Appl
18	528	68.1	151	33	US-60-444-371-2	Sequence 2, Appl
19	505	65.2	120	6	US-08-233-642A-55	Sequence 55, Appl
20	475	61.3	158	16	US-09-252-691C-5834	Sequence 5834, Ap
21	475	61.3	158	16	US-09-252-691C-5834	Sequence 5834, Ap
22	475	61.3	158	30	US-10-417-886-5834	Sequence 5834, Ap
23	471	60.8	109	19	US-09-543-407-34	Sequence 34, Appl
24	452	58.3	131	13	US-08-978-878-2	Sequence 2, Appl
25	452	58.3	131	21	US-09-741-873B-2	Sequence 2, Appl
26	348	44.9	109	19	US-09-543-407-35	Sequence 35, Appl
27	347	44.8	68	19	US-09-543-407-37	Sequence 37, Appl
28	248.5	32.1	70	19	US-09-543-407-32	Sequence 32, Appl
29	237	30.6	48	19	US-09-543-407-39	Sequence 39, Appl
30	111	14.3	445	29	US-10-369-493-20638	Sequence 20638, A
31	111	14.3	445	33	US-60-360-039-20638	Sequence 20638, A
32	107	13.8	186	16	US-09-252-691-5833	Sequence 5833, Ap
33	107	13.8	186	16	US-09-252-691C-5833	Sequence 5833, Ap
34	107	13.8	186	30	US-10-417-886-5833	Sequence 5833, Ap
35	98.5	12.7	151	19	US-09-543-407-6	Sequence 6, Appl
36	98	12.6	903	24	US-09-914-543-44	Sequence 44, Appl
37	96.5	12.5	151	19	US-09-543-407-8	Sequence 8, Appl
38	92.5	11.9	850	20	US-09-614-150A-24084	Sequence 24084, A
39	92.5	11.9	850	20	US-09-614-150A-24084	Sequence 24084, A
40	92.5	11.9	850	33	US-60-191-637-24179	Sequence 24179, A
41	92.5	11.9	850	33	US-60-191-681-19019	Sequence 19019, A
42	92.5	11.9	1028	20	US-09-614-150-14916	Sequence 14916, A
43	92.5	11.9	1028	20	US-09-614-150A-14916	Sequence 14916, A
44	92.5	11.9	1028	33	US-60-167-217-14978	Sequence 14978, A
45	92.5	11.9	1028	33	US-60-173-464-12194	Sequence 12194, A

ALIGNMENTS

RESULT 1
US-09-543-407-24
; Sequence 24, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b atgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.

US-09-543-407-24

Query Match 100.0%; Score 775; DB 19; Length 151;
Best Local Similarity 100.0%; Pred. No. 2.7e-74; Indels 0; Gaps 0;
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGGNGSGPDSSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGGNGSGPDSSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGGADNSTIETQNGFRNNATIDQWAKNYDQLVTRVVT 120
DB 61 SDARKSETTITQSGYNGADVGGADNSTIETQNGFRNNATIDQWAKNYDQLVTRVVT 120
QY 121 HEMAHANQATSDSSVMVRQVGFNNATANQY 151
DB 121 HEMAHANQATSDSSVMVRQVGFNNATANQY 151

RESULT 2

US-09-543-407-14
; Sequence 14, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-14

Query Match 92.1%; Score 714; DB 19; Length 151;
Best Local Similarity 91.1%; Pred. No. 9.2e-68; Indels 14; Gaps 2;
Matches 144; Conservative 0; Mismatches 0; Indels 14; Gaps 2;

QY 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGGNGSGPDSSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGGNGSGPDSSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGGADNSTIETQNGFRNNATIDQWAKNYDQLVTRVVT 113
DB 61 SDARKSETTITQSGYNGADVGGADNSTIETQNGFRNNATIDQWAKNYDQLVTRVVT 120
QY 114 LVTRVVTHEMAHANQATSDSSVMVRQVGFNNATANQY 151
DB 121 LVTRVVTHEMAHANQATSDSSVMVRQVGFNNATANQY 151

RESULT 3

US-09-543-407-30
; Sequence 30, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407

; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-30

Query Match 90.3%; Score 700; DB 19; Length 151;
Best Local Similarity 89.9%; Pred. No. 2.9e-66; Indels 16; Gaps 2;
Matches 143; Conservative 0; Mismatches 0; Indels 16; Gaps 2;

QY 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGGNGSGPDSSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGGNGSGPDSSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGGADNSTIETQNGFRNNATIDQWAKNYDQLVTRVVT 120
DB 61 SDARKSETTITQSGYNGADVGGADNSTIETQNGFRNNATIDQWAKNYDQLVTRVVT 112
QY 121 HEMAHANQATSDSSVMVRQVGFNNATANQY 151
DB 113 HEMAHANQATSDSSVMVRQVGFNNATANQY 151

RESULT 4

US-09-543-407-5
; Sequence 5, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Salmonella enteritidis
US-09-543-407-5

Query Match 89.0%; Score 690; DB 19; Length 151;
Best Local Similarity 91.4%; Pred. No. 3.4e-65; Indels 0; Gaps 0;
Matches 138; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGGNGSGPDSSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGGNGSGPDSSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGGADNSTIETQNGFRNNATIDQWAKNYDQLVTRVVT 120
DB 61 SDARKSETTITQSGYNGADVGGADNSTIETQNGFRNNATIDQWAKNYDQLVTRVVT 120
QY 121 HEMAHANQATSDSSVMVRQVGFNNATANQY 151
DB 121 NNAALVNQATSDSSVMVRQVGFNNATANQY 151

RESULT 5

US-08-233-642A-57
; Sequence 57, Application US/08233642A
; GENERAL INFORMATION:
; APPLICANT: Kay, William W.

APPLICANT: Collinson, S. Karen
APPLICANT: Clouthier, Sharon C.
APPLICANT: Doran, James L.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
BASED VACCINES
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed and Berry
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: U.S.A.
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233,642A
FILING DATE: 26-APR-1994
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: King, Joshua
REGISTRATION NUMBER: 35,570
REFERENCE/DOCKET NUMBER: 920043.403C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEEDANBERRY
INFORMATION FOR SEQ ID NO: 57:
SEQUENCE CHARACTERISTICS:
LENGTH: 151 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-233-642A-57

Query Match 88.9%; Score 689; DB 6; Length 151;
Best Local Similarity 91.4%; Pred. No. 4.4e-65;
Matches 138; Conservative 2; Mismatches 11; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALQ 60
QY 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNATIDQWNAKNDLVTRVVT 120
DB 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNATIDQWNAKNDLVTRVVT 120
QY 121 HEMAHANQASDSSVMVRQVGFNNATANQY 151
DB 121 NNPAALVQASDSSVMVRQVGFNNATANQY 151

RESULT 6
US-09-543-407-12
Sequence 12, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Collinson, S. Karen
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543,407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12
LENGTH: 151

TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-12

Query Match 84.5%; Score 653; DB 19; Length 151;
Best Local Similarity 79.8%; Pred. No. 1.9e-61;
Matches 134; Conservative 0; Mismatches 0; Indels 34; Gaps 2;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALQ 60
QY 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNATIDQWNAKNDLVTRVVT 109
DB 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNATIDQWNAKNDLVTRVVT 120
QY 110 -----NYDQLVTRVVTHEMAHANQASDSSVMVRQVGFNNATANQY 151
DB 121 NNAALVNYDQLVTRVVTHEMAHA-----NNATANQY 151

RESULT 7
US-09-543-407-22
Sequence 22, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543,407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 22
LENGTH: 151
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-22

Query Match 78.6%; Score 609; DB 19; Length 151;
Best Local Similarity 73.8%; Pred. No. 1.6e-56;
Matches 128; Conservative 0; Mismatches 0; Indels 46; Gaps 2;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALQ 60
QY 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNATIDQWNAKNDLVTRVVT 120
DB 61 SDARKSETTITQSGYNGADVGQAD-----NYDQLVTRVVT 97
QY 121 HEMAHANQASDSSVMVRQVGFNNATANQY 151
DB 98 HEMAHANQASDSSVMVRQVGFNNATANQY 151

RESULT 8
US-09-543-407-26
Sequence 26, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.

```
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-26

Query Match 78.6%; Score 609; DB 19; Length 151;
Best Local Similarity 82.1%; Pred. No. 1.6e-56;
Matches 124; Conservative 4; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGSGGPDSTLSIYQGSANAALALQ 60
DB 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGSGGPDSTLSIYQGSANAALYDQ 60
QY 61 SPARKSETTITQSGYNGADVQGGADNSTIELTQNGFRNNATIDQWNAKNDITVGYGG 120
DB 61 LVTRVVTTHMAHAGVNGADVQGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
QY 121 HEMAHANQATSSSVVMVRQVGFNNATANQY 151
DB 121 NNAALVNTQATSSSVVMVRQVGFNNATANQY 151

RESULT 9
US-09-543-407-18
; Sequence 18, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-18

Query Match 78.5%; Score 608; DB 19; Length 151;
Best Local Similarity 82.1%; Pred. No. 2.1e-56;
Matches 124; Conservative 5; Mismatches 22; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGSGGPDSTLSIYQGSANAALALQ 60
DB 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGSGGPDYDQVLTTRVVTTHMAHALQ 60
QY 61 SPARKSETTITQSGYNGADVQGGADNSTIELTQNGFRNNATIDQWNAKNDITVGYGG 120
DB 61 SPARKSETTITQSGYNGADVQGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
```

```
QY 121 HEMAHANQATSSSVVMVRQVGFNNATANQY 151
DB 121 NNAALVNTQATSSSVVMVRQVGFNNATANQY 151

RESULT 10
US-09-543-407-20
; Sequence 20, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-20

Query Match 78.2%; Score 606; DB 19; Length 151;
Best Local Similarity 82.1%; Pred. No. 3.4e-56;
Matches 124; Conservative 5; Mismatches 22; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGSGGPDSTLSIYQGSANAALALQ 60
DB 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGSGGPDSTLSIYQGSANAALALQ 60
QY 61 SPARKSETTITQSGYNGADVQGGADNSTIELTQNGFRNNATIDQWNAKNDITVGYGG 120
DB 61 SPARKYDQLVTRVVTTHMAHAGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
QY 121 HEMAHANQATSSSVVMVRQVGFNNATANQY 151
DB 121 NNAALVNTQATSSSVVMVRQVGFNNATANQY 151

RESULT 11
US-09-543-407-31
; Sequence 31, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Salmonella enteritidis
US-09-543-407-31

Query Match 77.8%; Score 603; DB 19; Length 131;
Best Local Similarity 90.1%; Pred. No. 5.9e-56;
Matches 118; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

QY 21 GVVPQWGGGNGHNGSGGPDSTLSIYQGSANAALALQSDARKSETTITQSGYNGAD 80
```

Db 1 GVVPQGGGNGHGGSSGPDSTLSIYQGSANAALALQSDARKSETTITQSGYNGAD 60
QY 81 VCGADNSTIETQNGFRNNATIDQWNAKNVDQLVTRVVTHEMAHANQATSDSSVMVRQV 140
Db 61 VCGADNSTIETQNGFRNNATIDQWNAKNSDITVGQYGGNNAALVNQATSDSSVMVRQV 120
QY 141 GFGNNATANQY 151
Db 121 GFGNNATANQY 131

RESULT 12

US-09-543-407-28
; Sequence 28, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-28

Query Match 77.4%; Score 600; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 1.5e-55;
Matches 123; Conservative 3; Mismatches 25; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHGGSSGPDSTLSIYQGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHGGSSGPDSTLSIYQGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVCGADNSTIETQNGFRNNATIDQWNAKNVDQLVTRVVT 120
Db 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHANQATSDSSVMVRQV 120
QY 121 HEMAHANQATSDSSVMVRQVFGNNATANQY 151
Db 121 NNAALVNQATSDSSVMVRQVFGNNATANQY 151

RESULT 13

US-09-543-407-16
; Sequence 16, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.405
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence

; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-16

Query Match 74.2%; Score 575; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 7.2e-53;
Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHGGSSGPDSTLSIYQGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVDQLVTRVVTHEMAHASGPDSTLSIYQGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVCGADNSTIETQNGFRNNATIDQWNAKNVDQLVTRVVT 120
Db 61 SDARKSETTITQSGYNGADVCGADNSTIETQNGFRNNATIDQWNAKNSDITVGQYGG 120
QY 121 HEMAHANQATSDSSVMVRQVFGNNATANQY 151
Db 121 NNAALVNQATSDSSVMVRQVFGNNATANQY 151

RESULT 14

US-09-543-407-7
; Sequence 7, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-543-407-7

Query Match 68.8%; Score 533; DB 19; Length 151;
Best Local Similarity 70.2%; Pred. No. 2.3e-48;
Matches 106; Conservative 18; Mismatches 27; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHGGSSGPDSTLSIYQGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQYGGGNGHGGSSGPNSELNIYQYGGNSALALQ 60
QY 61 SDARKSETTITQSGYNGADVCGADNSTIETQNGFRNNATIDQWNAKNVDQLVTRVVT 120
Db 61 TDARNSDLTITQGGGNGADVCGGSDSSIDLTDQRFNGSATLDQWNGKNSMTVKQFG 120
QY 121 HEMAHANQATSDSSVMVRQVFGNNATANQY 151
Db 121 GNGAIVDQIASNSVNVTVQFGNNATANQY 151

RESULT 15

US-08-978-878-4
; Sequence 4, Application US/08978878
; GENERAL INFORMATION:
; APPLICANT: NORMARK, Staffan
; APPLICANT: OLSEN, Arne
; TITLE OF INVENTION: FIBRONECTIN BINDING PROTEIN AS WELL AS ITS PREPARATION
; FILE REFERENCE: 012889-081
; CURRENT APPLICATION NUMBER: US/08/978,878
; CURRENT FILING DATE: 1997-11-26
; EARLIER APPLICATION NUMBER: SE 8801723-1
; EARLIER FILING DATE: 1998-05-06

```
; EARLIER APPLICATION NUMBER: US 07/347,189
; EARLIER FILING DATE: 1989-05-04
; EARLIER APPLICATION NUMBER: US 07/789,437
; EARLIER FILING DATE: 1991-11-06
; EARLIER APPLICATION NUMBER: US 07/970,846
; EARLIER FILING DATE: 1992-11-03
; EARLIER APPLICATION NUMBER: US 08/187,865
; EARLIER FILING DATE: 1994-01-28
; EARLIER APPLICATION NUMBER: US 08/318,519
; EARLIER FILING DATE: 1994-10-05
; EARLIER APPLICATION NUMBER: US 08/495,959
; EARLIER FILING DATE: 1995-06-28
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-08-978-878-4

Query Match      68.4%; Score 530; DB 13; Length 151;
Best Local Similarity 69.5%; Pred. No. 4.8e-48;
Matches 105; Conservative 19; Mismatches 27; Indels 0; Gaps 0;

QY      1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHGGNSGDPSTLSIYQGSANAALALQ 60
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB      1 MKLLKVAATAAIVFSGSAGVAVPQYGGGNGHGGNSGDPSTLSIYQGSANAALALQ 60
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY      61 SPARKSETTITQSGYNGADVGCGADNSTIELTQNGFRNNATIDOWNNAKNVDQLVTRVVT 120
        :|||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:
DB      61 TDARNSDLTITQGGGNGADVGCGSDSDSIDLTQGFNGSATLDQWNGKNSMTVKQFEG 120
        :|||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:

QY      121 HEMAHANQTASDSVNVQVGFNNATANQY 151
        :|||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:
DB      121 GNCAAVDQTASNSVNVTVQVGFNNATAHQY 151
        :|||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:
```

Search completed: August 2, 2004, 15:26:45
Job time: 167.9 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:49:38 ; Search time 17.8 seconds
(without alignments)
888.146 Million cell updates/sec

Title: US-09-543-407-24
Perfect score: 775
Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVFGNNATANYQ 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 601315 seqs, 104695340 residues

Total number of hits satisfying chosen parameters: 601315

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents AA New:

1: /cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pep.*
2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep.*
3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep.*
4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep.*
5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep.*
6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep.*
7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	530	68.4	151	5	US-09-741-873C-4
2	452	58.3	131	5	US-09-741-873C-2
3	91	11.7	719	5	US-09-248-796A-17559
4	90.5	11.7	258	6	US-10-425-115-300390
5	90.5	11.7	295	6	US-10-425-115-312468
6	90.5	11.7	299	6	US-10-170-205E-35751
7	90.5	11.7	520	6	US-10-479-638-21
8	90.5	11.7	1203	6	US-10-170-205E-741
9	89.5	11.7	1327	1	PCT-US04-02338-49
10	89.5	11.5	256	6	US-10-425-115-301334
11	85.5	11.0	388	5	US-09-248-796A-17306
12	85	11.0	535	1	PCT-US04-21492-88
13	84	10.8	376	6	US-10-491-733-2
14	83.5	10.8	179	6	US-10-425-115-346132
15	83.5	10.8	573	7	US-60-565-632-7907
16	83.5	10.8	573	7	US-60-579-062-7907
17	83	10.7	889	5	US-09-952-267B-15
18	83	10.7	889	5	US-10-872-768-15
19	83	10.7	889	6	US-10-872-769-15
20	81.5	10.5	508	6	US-10-425-115-285216
21	81.5	10.5	532	6	US-10-425-115-285214
22	81	10.5	131	7	US-60-565-632-11109
23	81	10.5	131	7	US-60-579-062-11109
24	81	10.5	201	6	US-10-425-115-309662
25	81	10.5	376	6	US-10-425-115-342526
26	81	10.5	382	6	US-10-771-241-299

27 81 10.5 659 1 PCT-US04-09385-4 Sequence 4, Appli
28 81 10.5 659 6 US-10-809-790-4 Sequence 4, Appli
29 80.5 10.4 246 6 US-10-854-439-511 Sequence 511 App
30 80.5 10.4 389 6 US-10-724-972A-4638 Sequence 4638, Ap
31 80.5 10.4 891 7 US-60-566-425-574 Sequence 574, App
32 80.5 10.4 891 7 US-60-576-812-606 Sequence 606, App
33 80.5 10.4 894 7 US-60-566-425-570 Sequence 570, App
34 80.5 10.4 894 7 US-60-566-425-571 Sequence 571, App
35 80.5 10.4 894 7 US-60-566-425-575 Sequence 575, App
36 80.5 10.4 894 7 US-60-566-425-578 Sequence 578, App
37 80.5 10.4 894 7 US-60-576-812-602 Sequence 602, App
38 80.5 10.4 894 7 US-60-576-812-603 Sequence 603, App
39 80.5 10.4 894 7 US-60-576-812-607 Sequence 607, App
40 80.5 10.4 894 7 US-60-576-812-610 Sequence 610, App
41 80.5 10.4 898 6 US-10-170-205E-12329 Sequence 12329, A
42 80.5 10.4 898 7 US-60-566-425-577 Sequence 577, App
43 80.5 10.4 898 7 US-60-576-812-609 Sequence 609, App
44 79 10.2 132 6 US-10-425-115-351875 Sequence 351875, A
45 79 10.2 218 5 US-09-248-796A-14423 Sequence 14423, A

ALIGNMENTS

RESULT 1

US-09-741-873C-4
; Sequence 4; Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012899-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873C-4

Query Match 68.4%; Score 530; DB 5; Length 151;

Best Local Similarity 69.8%; Pred.No. 1.8e-39;
Matches 105; Conservative 19; Mismatches 27; Indels 0; Gaps 0;

Qy 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQVGSANAALQ 60

Db 1 MKLLKVAIAAIVFSGSAVAGVVPQYGGGNGHNGGNSGPNSELNIYQYGGGNSALALQ 60

Qy 61 SDARKSETTITQSGYNGADVGQGANSTIELTQNGFRNNATIDQNAKNDLVTRVVT 120

Db 61 TDARNEDLITTHGGNGADVGQGSDDSLDLTQRFNGSATLDQNGKNSMTVQFGG 120

Qy 121 HEMAHANQATSDSSVMVRQVFGNNATANYQ 151

Db 121 GNGAVDQATASNSVNVTVQFGNNATAHOY 151

```

RESULT 2
US-09-741-873C-2
; Sequence 2, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Norman, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873C-2

Query Match          58.3%; Score 452; DB 5; Length 131;
Best Local Similarity 66.4%; Pred. No. 1.2e-32;
Matches      87; Conservative 18; Mismatches 26; Indels 0; Gaps 0;

QY      21  GVVPQGGGNNHGGNGSGPDTLSTIYQYGSANAALALQSDARKSETTITQSGYNGAD 80
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db       1  GVVPYQGGGNNHGGGNGSGPNSLNTYQYGGNSALALQTDARNSDLTTTQHGNGAD 60
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY      81  VGGGANSTIELTQNGFRNNATIDQNAKNDQLVTFVVTHEMAHANCTASDSSVMVROY 140
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db       61  VGGGSDSSLDLITQRFGNSATLDQNGKNSMTVTKQFGGNGAAVDQTASNSVNTQV 120
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY      141  GFGNNATANQY 151
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db       121  GFGNNATAHQY 131
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT 3
US-09-248-796A-17559
; Sequence 17559, Application US/09248796A
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 17559
; LENGTH: 719
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-17559

Query Match          11.7%; Score 91; DB 5; Length 719;
Best Local Similarity 23.7%; Pred. No. 6.2;
Matches      28; Conservative 25; Mismatches 59; Indels 6; Gaps 2;

```

Matches 29; Conservative 9; Mismatches 44; Indels 5; Gaps 2;
QY 26 WGGGNNHNGGNSGPDSTLSIYQGSANAALALQSDARKSETTITQSGYNGADVQCGA 85
Db 138 YGGGYSGGGYSYG-GYAANGYGVGSGGNYSNASGGYSGS---DGYNGAASGGYA 192
QY 86 DNSITELTQNGFRNNAIDQWNAKND 112
Db 193 NNLSSGYSNGRYNTIGSSDNGTGGYN 219

RESULT 6
US-10-170-205E-35751
; Sequence 35751, Application US/10170205E
; GENERAL INFORMATION:
; APPLICANT: ADAMS, Mark
; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN
; TITLE OF INVENTION: CAPTURE AGENTS, AND USES THEREOF
; FILE REFERENCE: CL001381
; CURRENT APPLICATION NUMBER: US/10/170,205E
; CURRENT FILING DATE: 2002-06-13
; NUMBER OF SEQ ID NOS: 40312
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 35751
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-170-205E-35751

Query Match 11.7%; Score 90.5; DB 6; Length 299;
Best Local Similarity 28.4%; Pred. No. 2.5;
Matches 42; Conservative 16; Mismatches 61; Indels 29; Gaps 6;
QY 6 VAAFAAI-VVSGALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANAALALQSDAR 64
Db 99 VAAAPVPAVSTSSAAGVAEPNAGSGNNSSPSSTSS-SSSSPSPGSSLAESPEAA 157
QY 65 KSETTIT---QSGYNGADVQCGADNSTIELTQNGFRNNAIDQWNAKNDQLVTRVVT 120
Db 158 GVSSTAPLPGGAAGTGTGVPVSGALRELLEACRNG-----DVSrvKRLVDA--- 204
QY 121 HEMAHANQTASD-----SSVMVROVGF 143
Db 205 -----ANVNAKDMAGRKSPPLHFAAGFG 227

RESULT 7
US-10-479-638-21
; Sequence 21, Application US/10479638
; GENERAL INFORMATION:
; APPLICANT: Don A. Roth
; APPLICANT: Randolph V. Lewis
; TITLE OF INVENTION: Expression of Spider Silk Proteins in Higher Plants
; FILE REFERENCE: WFO/02-0004US
; CURRENT APPLICATION NUMBER: US/10/479,638
; CURRENT FILING DATE: 2003-12-03
; PRIOR APPLICATION NUMBER: PCT/US02/18256
; PRIOR FILING DATE: 2002-06-06
; PRIOR APPLICATION NUMBER: 60/296,184
; PRIOR FILING DATE: 2001-06-06
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 21
; LENGTH: 520
; TYPE: PRT
; ORGANISM: Argiope trifasciata
US-10-479-638-21

Query Match 11.7%; Score 90.5; DB 6; Length 520;
Best Local Similarity 24.6%; Pred. No. 4.8;
Matches 34; Conservative 27; Mismatches 64; Indels 13; Gaps 4;

QY 15 SGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANAALALQSDARKSET-TITQS 73
Db 347 AGAGAAAAASAGAGAGYGGYGVAGSS-----ISYGTSSATSSSTASSRSGIVTSG 402
QY 74 GYNGADVQCGADNSTIELTQNGFRNNAIDQWNAKNDQLVTRVVTHEMAHANQTASDS 133
Db 403 GYGAGAAAGAGAGAAAGAGSYG-----SISRLLSA---EAVNRVSSNIGAVASGGASAL 454
QY 134 SVMVROVGFNNATANQY 151
Db 455 PGVISNIFSGVSSAGSY 472

RESULT 8
US-10-170-205E-741
; Sequence 741, Application US/10170205E
; GENERAL INFORMATION:
; APPLICANT: ADAMS, Mark
; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN
; TITLE OF INVENTION: CAPTURE AGENTS, AND USES THEREOF
; FILE REFERENCE: CL001381
; CURRENT APPLICATION NUMBER: US/10/170,205E
; CURRENT FILING DATE: 2002-06-13
; NUMBER OF SEQ ID NOS: 40312
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 741
; LENGTH: 1203
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-170-205E-741

Query Match 11.7%; Score 90.5; DB 6; Length 1203;
Best Local Similarity 28.4%; Pred. No. 13;
Matches 42; Conservative 16; Mismatches 61; Indels 29; Gaps 6;
QY 6 VAAFAAI-VVSGALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANAALALQSDAR 64
Db 99 VAAAPVPAVSTSSAAGVAEPNAGSGNNSSPSSTSS-SSSSPSPGSSLAESPEAA 157
QY 65 KSETTIT---QSGYNGADVQCGADNSTIELTQNGFRNNAIDQWNAKNDQLVTRVVT 120
Db 158 GVSSTAPLPGGAAGTGTGVPVSGALRELLEACRNG-----DVSrvKRLVDA--- 204
QY 121 HEMAHANQTASD-----SSVMVROVGF 143
Db 205 -----ANVNAKDMAGRKSPPLHFAAGFG 227

RESULT 9
PCT-US04-02338-49
; Sequence 49, Application PC/TUS0402338
; GENERAL INFORMATION:
; APPLICANT: EXELIXIS, INC.
; TITLE OF INVENTION: MAPKs AS MODIFIERS OF THE APC AND AXIN PATHWAYS AND METHODS OF
; TITLE OF INVENTION: USE
; FILE REFERENCE: EX04-003C-PC
; CURRENT APPLICATION NUMBER: PCT/US04/02338
; CURRENT FILING DATE: 2004-01-28
; PRIOR APPLICATION NUMBER: US60/443,484
; PRIOR FILING DATE: 2003-01-29
; PRIOR APPLICATION NUMBER: US60/447,358
; PRIOR FILING DATE: 2003-02-11
; PRIOR APPLICATION NUMBER: US60/461,789
; PRIOR FILING DATE: 2003-04-10
; PRIOR APPLICATION NUMBER: US60/470,684
; PRIOR FILING DATE: 2003-05-14
; PRIOR APPLICATION NUMBER: US60/479,650
; PRIOR FILING DATE: 2003-06-19
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 49
; LENGTH: 1327
; TYPE: PRT

APPLICANT: Weislo, Laura J.
TITLE OF INVENTION: Nucleic Acid Molecules from Rice Encoding Proteins for Abiotic St
FILE OF INVENTION: Resistance, Yield, Disease Resistance and Nutritional Quality an
FILE REFERENCE: 60127WOPCT
CURRENT APPLICATION NUMBER: US/10/491,733
CURRENT FILING DATE: 2004-04-05
PRIOR APPLICATION NUMBER: 60/334,501
PRIOR FILING DATE: 2001-11-30
NUMBER OF SEQ ID NOS: 64
SOFTWARE: Patent in version 3.1
SEQ ID NO 2
LENGTH: 376
TYPE: PRT
ORGANISM: Oryza sativa
US-10-491-733-2

Query Match 10.8%; Score 84; DB 6; Length 376;
Best Local Similarity 31.5%; Pred. No. 12;
Matches 34; Conservative 15; Mismatches 39; Indels 20; Gaps 6;

QY 17 SALAGVVPQWGGGNNHGGSSGPDSTLSIVQGSANAALALOS--DARKSETTITQSG 74
DB 50 SALAG-----GGGG--GGGGGFGKD-----FGSMMDLLSIWTAEESQAVASAG 95

QY 75 YNGADVGGADNSTIELTONGFRNNATIDOWNAKNYDQLVTRVVTHE 122
DB 96 SAAGVGVAVGAPPTSLQ-ROGSLTLPRTL---SAKTVDVVRNLVRDE 139

RESULT 14
US-10-425-115-346132
Sequence 346132, Application US/10425115
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J.
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihue
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
FILE OF INVENTION: Plants
FILE REFERENCE: 38-21(53222)B
CURRENT APPLICATION NUMBER: US/10/425,115
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 369326
SEQ ID NO 346132
LENGTH: 179
TYPE: PRT
ORGANISM: Zea mays
FEATURE:
OTHER INFORMATION: Clone ID: MRT4577_78839C.1.pap
US-10-425-115-346132

Query Match 10.8%; Score 83.5; DB 6; Length 179;
Best Local Similarity 30.4%; Pred. No. 5.8;
Matches 28; Conservative 8; Mismatches 33; Indels 23; Gaps 4;

QY 19 LAGVFPQWGGG-----NHNGGSSGPDSTLSIVQGSANAALALQSDARKSETTI 70
DB 64 LDGLLGLGGGGGLDGLLGLTGGGGGKQENQADSGNAOEGSGNA-----QEDSGNAQ--- 116

QY 71 TOSGYNGADVGGADNSTIELTONGFRNNAT 102
DB 117 EESGNNQAGAGAGA-----ENGGAANNGT 140

RESULT 15
US-60-565-632-7907
Sequence 7907, Application US/60565632
GENERAL INFORMATION:
APPLICANT: Monsanto Technology, LLC
APPLICANT: Baum, James A
APPLICANT: Kovalic, David K.
APPLICANT: Larosa, Thomas J.
APPLICANT: Lu, Maolong

APPLICANT: Munyikwa, Tichifa R. I.
APPLICANT: Roberts, James K.
APPLICANT: Wu, Wei
APPLICANT: Zhang, Bei
TITLE OF INVENTION: Methods for Genetic Control of Insect Infestations in Plants and
FILE OF INVENTION: Compositions Thereof
FILE REFERENCE: 38-21(53403)B
CURRENT APPLICATION NUMBER: US/60/565,632
CURRENT FILING DATE: 2004-04-27
NUMBER OF SEQ ID NOS: 15449
SOFTWARE: Patent in version 3.2
SEQ ID NO 7907
LENGTH: 573
TYPE: PRT
ORGANISM: Diabrotica virgifera
US-60-565-632-7907

Query Match 10.8%; Score 83.5; DB 7; Length 573;
Best Local Similarity 27.8%; Pred. No. 22;
Matches 45; Conservative 15; Mismatches 73; Indels 29; Gaps 7;

QY 10 AAIWVSGSALAGVVPQWGGGNNHGGSSGPDSTLSIVQGSANAALA-----LQSD 62
DB 155 AAADNNGSADAA-----QGNDNRAAENANANADAQTDAQA-QSANEANAENANADAQND 208

QY 63 ARKSETTITQSGYNGADVGGADN-STIELTGN-----GFRNNATID-----QWNAX 109
DB 209 AAOANDNGAAEAENNGNADAAQGTDEAAEENSGNENGTGAENNANADAQTDDVAQGSTNEA 268

QY 110 NYDQLVTRVVTHEMALANQ---TASDSSVMVRQVGFNNATA 148
DB 269 NAENNANADVQNDAAQAQANENGAAEENSGNADAAQGTDENGAAA 310

Search completed: August 2, 2004, 15:29:54
Job time : 17.8 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:39:53 ; Search time 9.4 Seconds
(without alignments)
1545.204 Million cell updates/sec

Title: US-09-543-407-24

Perfect score: 775
Sequence: 1 MLLKVAFAAIVVSGSALA.....DSSVMVRQVGFNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

PIR_78:*
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	690	89.0	151	2 JC6039	fimbrin protein ag
2	690	89.0	151	2 A10635	major curlin chain
3	533	68.8	151	2 S70788	curlin protein csg
4	511.5	66.0	152	2 D90806	curlin major subun
5	511.5	66.0	152	2 H85665	hypothetical prote
6	101	13.0	409	2 T20847	hypothetical prote
7	99	12.8	1748	2 S42136	cnjB protein - Tet
8	98.5	12.7	151	2 JC6040	fimbrin protein ag
9	98.5	12.7	151	2 AH0635	nucleation compone
10	96.5	12.5	151	2 S70787	curlin nuclearor p
11	96.5	12.5	151	2 C90806	minor curlin subun
12	96.5	12.5	151	2 G85665	curlin minor chain
13	93.5	12.1	552	2 T70604	probable PPE prote
14	92.5	11.9	440	2 AD1539	probable sugar ABC
15	92.5	11.9	1028	2 A56038	DNA-binding protei
16	92.5	11.9	1213	2 S16356	ovo protein - frui
17	90	11.6	145	2 AD3143	conserved hypothet
18	90	11.6	145	2 H98144	hypothetical prote
19	90	11.6	347	2 B39112	merozoite 45K surf
20	89.5	11.5	256	2 T03371	glycine-rich prote
21	89	11.5	262	2 S00275	tail fiber protein
22	89	11.5	573	2 C86266	F3f19.21 protein -
23	89	11.5	2174	2 E95965	hypothetical glyci
24	88.5	11.4	221	2 A47369	RNA-binding protei
25	87	11.2	590	1 A45621	leishmanolysin (EC
26	87	11.2	599	2 B42049	leishmanolysin (EC
27	87	11.2	599	2 A44951	leishmanolysin (EC
28	87	11.2	602	1 PL0221	leishmanolysin (EC
29	87	11.2	646	1 S19916	leishmanolysin (EC

RESULT 1

JC6039
fimbrin protein agfa precursor - Salmonella enteritidis
C:Species: Salmonella enteritidis
C:Date: 31-Dec-1996 #sequence revision 31-Dec-1996 #text_change 08-Oct-1999
C:Accession: JC6039; PC6015; A44898
R:Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Banser, P.A.; Kay, W.W.
J. Bacteriol. 178, 662-667, 1996
A:Title: Salmonella enteritidis agfBAC operon encoding thin, aggregative fimbrinae.
A:Reference number: JC6039; MUID:96146512; PMID:8550497
A:Accession: JC6039
A:Molecule type: DNA
A:Residues: 1-151 <COL>
A:Cross-references: GB:U43280; NID:g1184712; PIDN:AAC43599.1; PID:g1184714
A:Accession: PC6015
A:Molecule type: protein
A:Residues: 21-52 <CO2>
A:Experimental source: strain 27655-3b
A:Note: the authors translated the codon ACG for residue 44 as Ile
R:Collinson, S.K.; Emody, L.; Muller, K.H.; Trust, T.J.; Kay, W.W.
J. Bacteriol. 173, 4773-4781, 1991
A:Title: Purification and characterization of thin, aggregative fimbrinae from Salmonella
A:Reference number: A44898; MUID:9310586; PMID:1677357
A:Contents: 27655
A:Accession: A44898
A>Status: preliminary
A:Molecule type: protein
A:Residues: 21-33 <CO3>
A:Note: sequence extracted from NCBI backbone (NCBIP:45936)
C:Genetics:
A:Gene: agfa
C:Function:
A:Description: major component of thin aggregative fimbrinae
A:Note: fimbrinae bind to fibronectin, plasminogen, tissue plasminogen activator
C:Keywords: fimbrina
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-151/Product: fimbrin protein agfa #status experimental <MAT>

Query Match 89.0%; Score 690; DB 2; Length 151;
Best Local Similarity 91.4%; Pred. No. 4.7e-51;
Matches 138; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

QY 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSSGPDSTLSIYQVGSANAALQ 60
DB 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSSGPDSTLSIYQVGSANAALQ 60
QY 61 SDARKSETTTTQSGYGNAGDVGGGADNSTTIELTQNGFRNNATIDQWNAKYDQAVTRVWT 120
DB 61 SDARKSETTTTQSGYGNAGDVGGGADNSTTIELTQNGFRNNATIDQWNAKYDQAVTRVWT 120
QY 121 HEMAFANQTASDSSVMVRQVGFNNATANQY 151

CT147 hypothetical
conserved hypothet
probable disease r
probable PPE prote
probable secreted
protein C05B5.3 [1
protein kinase sgg
hypothetical prote
hypothetical prote
probable sugar ABC
hypothetical prote
hemolysin (import
leishmanolysin (EC
hypothetical prote
probable outer mem

ALIGNMENTS

30 87 11.2 1537 2 F86509
31 87 11.2 1537 2 C81558
32 86 11.1 447 2 G84687
33 86 11.1 582 2 F70675
34 85.5 11.0 438 2 T35789
35 85 11.0 401 2 C88571
36 85 11.0 575 2 S35327
37 85 11.0 967 2 S66852
38 84.5 10.9 340 2 A83401
39 84 10.8 439 2 AC1182
40 84 10.8 764 2 H71607
41 83.5 10.8 423 2 T19581
42 83.5 10.8 1635 2 AI0452
43 83 10.7 639 2 C42049
44 82.5 10.6 343 2 T05221
45 82 10.6 382 2 F90892

Db 121 NNAALVNOTASDSSVMVRQVGFNNATANOY 151

RESULT 2

AI0635
major curlin chain precursor [imported] - Salmonella enterica subsp. enterica serovar Typhimurium
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AI0635
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Conerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AI0635

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-151 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD08268.1; PID:g16502315; GSPDB:GN00176

C:Genetics:

A:Gene: STY181

Query Match 89.0%; Score 690; DB 2; Length 151;
Best Local Similarity 91.4%; Pred. No. 4.7e-51;
Matches 138; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60

DB 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGADVGOGADNSTIETQNGFRNATIDQWNAKNDQLVTRVVT 120

DB 61 SDARKSETTITQSGYNGADVGOGADNSTIETQNGFRNATIDQWNAKNDQLVTRVVT 120

QY 121 HEMAHANQASDSSVMVRQVGFNNATANOY 151

DB 121 NNAALVNOTASDSSVMVRQVGFNNATANOY 151

RESULT 3

S70788
curlin protein csgA precursor - Escherichia coli (strain K-12)
N:Alternate names: csgA protein; major curlin protein
C:Species: Escherichia coli
C:Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 01-Mar-2002
C:Accession: S70788; G64846; S31202; S34559
R:Hammar, M.; Arngqvist, A.; Bian, Z.; Olsen, A.; Normark, S.
Mol. Microbiol. 18, 661-670, 1995

A:Title: Expression of two csg operons is required for production of fibronectin- and Co
A:Reference number: S70788; MUID:96414468; PMID:8817489
A:Accession: S70788

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-151 <HAM>

A:Cross-references: EMBL:X90754; NID:g1147558; PIDN:CAAG2282.1; PID:g1147564

A:Experimental source: strain K12, substrain W3110

A:Note: The nucleotide sequence was submitted to the EMBL Data Library, August 1995
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A6720; MUID:97426617; PMID:9278503

A:Accession: G64846

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-151 <BLAT>

A:Cross-references: GB:AE000205; GB:U00096; NID:g1787265; PIDN:AAC74126.1; PID:g1787279;

A:Experimental source: strain K-12, substrain MG1655

R:Olsen, A.; Arngqvist, A.; Hammar, M.; Sukupolvi, S.; Normark, S.
Mol. Microbiol. 7, 523-536, 1993

A:Title: The Rpos sigma factor relieves H-NS-mediated transcriptional repression of csgA
A:Reference number: S31202; MUID:93211294; PMID:8459772
A:Accession: S31202

A:Molecule type: DNA

A:Residues: 1-6,'V',8-151 <OLS1>

A:Cross-references: EMBL:L04979

A:Accession: S34560

A:Molecule type: protein

A:Residues: 21-42;44-50 <OLS2>

R:Olsen, A.N.; Arngqvist, A.M.

A:Reference number: S34559

A:Accession: S34559

A:Molecule type: DNA

A:Residues: 1-133,'RDRDSGLW' <OLS3>

A:Cross-references: EMBL:L04979; NID:g290424; PIDN:AAA23616.1; PID:g290425

A:Experimental source: strain K-12, substrain W3110

C:Genetics:

A:Gene: csgA

A:Map position: 23.15

C:Function:

A:Description: major component of wild-type curli; interaction between CsgA and CsgB tri
A:Note: curli are thin, coiled fibers expressed on the surface of Escherichia coli that
and H-kininogen; in the absence of CsgA, CsgB can self-assemble into polymers
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-151/Product: curlin #status experimental <MAT>

Query Match 68.8%; Score 533; DB 2; Length 151;
Best Local Similarity 70.2%; Pred. No. 7.1e-38;
Matches 106; Conservative 18; Mismatches 27; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60

DB 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGADVGOGADNSTIETQNGFRNATIDQWNAKNDQLVTRVVT 120

DB 61 TDARNSDLTITQHGNGGNGADVGOGSDSSIDLTFQGFNSATLDQWNGKNSMTVKQFGG 120

QY 121 HEMAHANQASDSSVMVRQVGFNNATANOY 151

DB 121 GNGHVDQATASNSVNVTVQVGFNNATANOY 151

RESULT 4

D90806
curlin major subunit CsgA [imported] - Escherichia coli (strain O157:H7, substrain RIMD
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: D90806
R:Hayaishi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: D90806

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-152 <HAY>

A:Cross-references: GB:BA000007; PIDN:BA834843.1; PID:g13360880; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RIMD 050952

C:Genetics:

A:Gene: ECs1420

Query Match 66.0%; Score 511.5; DB 2; Length 152;

Best Local Similarity 68.4%; Pred. No. 4.6e-36;

Matches 104; Conservative 19; Mismatches 28; Indels 1; Gaps 1;

QY 1 MKLLKVAAPAAIIVVSGSALAGVVPQW-GGGGNGHNGGNSGPDSTLSIYQYGSANAALAL 59

DB 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGHNGGNSGPNSELNIYQYGGNSALAL 60

QY 60 QSDARKSETTITQSGYNGADVGQGDADNSTIETQNGFRNATIDQWNAKNDQLVTRVVT 119


```
Db 61 QADAENSLTITQHGGNGADVGGSDSSIDLTRQFGNSATLDQWNGKDSHTVTKQFG 120
QY 120 THEMAHANOTASDSSVMVRQVFGNNATANQY 151
Db 121 GGNGAAMDQATASNTVNTVQVFGNNATAHQY 152

RESULT 5
H85665
hypothetical protein csgA [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
C:Species: Escherichia coli
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: H85665
R:Perina, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: H85665
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-152 <STO>
A:Cross-references: GB:AE005174; NID:G12514574; PIDN:AA655788.1; GSPDB:GN00145; UWGP:Z16
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: csgA

Query Match 66.0%; Score 511.5; DB 2; Length 152;
Best Local Similarity 68.4%; Pred. No. 4.6e-36;
Matches 104; Conservative 19; Mismatches 26; Indels 1; Gaps 1;

QY 1 MKLLKVAFAATVVGSSALAGVVPQW-GGGGNHGGNSGPDSTLSIYQYGSANAALAL 59
Db 1 MKLLKVAFAATVVGSSALAGVVPQYGGGGNHGGNSGPNSELNIIYQYGGNSALAL 60

QY 60 QSDARKSETTITQSGYNGADVQGGADNSTIELTQNGFRNATIDQWNAKNDQLVTRV 119
Db 61 QADAENSLTITQHGGNGADVGGSDSSIDLTRQFGNSATLDQWNGKDSHTVTKQFG 120

QY 120 THEMAHANOTASDSSVMVRQVFGNNATANQY 151
Db 121 GGNGAAMDQATASNTVNTVQVFGNNATAHQY 152

RESULT 6
T20847
hypothetical protein Fl3E9.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 02-Aug-2002
C:Accession: T20847
R:McMurray, A.
submitted to the EMBL Data Library, February 1996
A:Reference number: Z19332
A:Accession: T20847
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-409 <WIL>
A:Cross-references: EMBL:Z69383; PIDN:CAA93412.1; GSPDB:GN00022; CESP:Fl3E9.4
A:Experimental source: clone Fl3E9
C:Genetics:
A:Gene: CESP:Fl3E9.4
A:Map position: 4
A:Introns: 32/1; 275/3; 337/3
C:Superfamily: loricrin

Query Match 13.0%; Score 101; DB 2; Length 409;
Best Local Similarity 23.3%; Pred. No. 0.4;
Matches 37; Conservative 26; Mismatches 66; Indels 30; Gaps 5;

QY 15 SGLSALAGVVPQWGGGNHGGNSGPDSTLSIYQYGSANAALALQSDARKSETTITQSG 74
Db 134 SQQAGSNNVSGGGQGGYQNGFGGQSGFGSGWGSNLSNNSGNNQGG--SSSG 191
```

```
QY 75 YGN--GADVQGGADNSTIE---LTQNGFRNN-----ATIDQWNAK 109
Db 192 YQNNQGHQCGGGGSHSSSVMSNNGSYGNNNGTPTPSFLNNVSSAAQDYINIV 251
QY 110 NYDQLVTRVTHEMAHANOTASDSSVMVRQVQVGGNNATA 148
Db 252 NKKSLTTNQINEQ---ASNWASANSVQAQYIQYETNRS 287
```

```
RESULT 7
S42136
cnjB protein - Tetrahymena thermophila
C:Species: Tetrahymena thermophila
C>Date: 19-Mar-1997 #sequence_revision 18-Jul-1997 #text_change 07-Dec-1999
C:Accession: S42136; S42135; S03650
R:Taylor, F.M.; Martindale, D.W.
submitted to the EMBL Data Library, October 1992
A:Reference number: S42136
A:Accession: S42136
A:Molecule type: DNA
A:Residues: 1-1748 <TAY>
A:Cross-references: EMBL:L03710; NID:G161751; PID:G161752
R:Taylor, F.M.; Martindale, D.W.
Nucleic Acids Res. 21, 4610-4614, 1993
A:Title: Retroviral-type zinc fingers and glycine-rich repeats in a protein encoded by
A:Reference number: S42135; MUID:94051569; PMID:8233798
A:Accession: S42135
A:Molecule type: DNA
A:Residues: 1164-1174; 1179-1198; 1233-1252; 1285-1293; 1297-1309; 1316-1326; 1331-1341; 1343-
A:Cross-references: EMBL:L03710
R:Martindale, D.W.; Taylor, F.M.
Nucleic Acids Res. 16, 2189-2201, 1988
A:Title: Multiple introns in a conjugation-specific gene from Tetrahymena thermophila.
A:Reference number: S03650; MUID:88189811; PMID:3357771
A:Accession: S03650
A:Molecule type: DNA
A:Residues: 236-250; '1', 252-255, 'N', 257-773 <MAR>
A:Cross-references: EMBL:X06462
C:Genetics:
A:Gene: cnjB
A:Genetic code: SGC5
A:Introns: 85/3; 136/1; 157/3; 201/2; 290/2; 327/3; 499/1; 573/2; 607/3; 708/3; 777/3;
C:Keywords: zinc finger
F:1164-1450/Region: Glycine-rich
F:1451-1464/Region: zinc finger CCHC motif
F:1478-1491/Region: zinc finger CCHC motif
F:1501-1514/Region: zinc finger CCHC motif
F:1530-1543/Region: zinc finger CCHC motif
F:1555-1568/Region: zinc finger CCHC motif
F:1579-1592/Region: zinc finger CCHC motif
F:1602-1615/Region: zinc finger CCHC motif
F:1626-1748/Region: glycine-rich

Query Match 12.8%; Score 99; DB 2; Length 1748;
Best Local Similarity 28.0%; Pred. No. 3.1;
Matches 37; Conservative 16; Mismatches 31; Indels 48; Gaps 6;

QY 25 QWGGGNGHGG---GNSSGPDSTLSIYQYGSANAALALQSDARKSETTIT---QSGYGN 77
Db 1640 QFGGGGNSGQSGWGTSSGSDWN-----CQSNVQESITTSGGGWSGSGN 1685

QY 78 GADVQGGADNSTIELTQNGFRNATIDQWNAKNDQLVTRVTHEMAHANOTASDSSVMV 137
Db 1686 QTGGGWSNDN-----QQQNTGGGGGSSN-----SNQTNSS--- 1722

QY 138 QVGFQGNATAN 149
Db 1723 ----WGSNNQAS 1730
```

```
RESULT 8
JC6040
```

fimbria protein agfB precursor - Salmonella enteritidis

C:Species: Salmonella enteritidis
C>Date: 31-Dec-1996 #sequence_revision 31-Dec-1996 #text_change 08-Oct-1999
C:Accession: J06040
R:Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Bansen, P.A.; Kay, W.W.
J. Bacteriol. 178, 662-667, 1996
A:Title: Salmonella enteritidis agfBAC operon encoding thin, aggregative fimbriae.
A:Reference number: JC6039; MUID:96146512; PMID:8550497
A:Accession: J06040
A:Molecule type: DNA
A:Residues: 1-151 <COL>
A:Cross-references: GB:U43280; NID:G1184712; PID:AA43598.1; PID:G1184713
A:Experimental source: strain 276755-3b
C:Genetics:
A:Gene: agfB
C:Function:
A:Description: minor component of thin aggregative fimbriae
A:Note: fimbriae bind to fibronectin, plasminogen, tissue plasminogen activator
C:Keywords: fimbria
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-151/Product: fimbria protein agfB #status predicted <MAT>

Query Match 12.7%; Score 98.5; DB 2; Length 151;
Best Local Similarity 28.8%; Pred. No. 0.21;
Matches 34; Conservative 17; Mismatches 48; Indels 19; Gaps 5;

QY 51 GSANAALALQSDARKSE---TTITQSGYNGADVQ-GADNST-----IELTQ 94

Db 14 GAGCIATATNYDLARSEYNAFVNELSKSFNQAAIIGVGTDSARVRQEGSKLLSVISQ 73

QY 95 NGRFNATIDQNAKND-OLTRVVTTHMAHANTASDSSVMVQVGFNNATANY 151

Db 74 EGGNNRAKVDQ--AGNYPAYIEQTGNANDASISQAYGNSAAIIQKSGNKNATITQY 129

RESULT 9

AH0635

nucleation component of curlin monomers [imported] - Salmonella enterica subsp. enterica

C:Species: Salmonella enterica subsp. enterica serovar Typhi

A:Note: this species has also been called Salmonella typhi

C>Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002

C:Accession: AH0635

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moulé, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.

A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar

A:Reference number: AB0502; MUID:21534947; PMID:11677608

A:Accession: AH0635

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-151 <PAR>

A:Cross-references: GB:AL513382; PID:CAD08267.1; PID:G16502314; GSPDB:GN00176

C:Genetics:

A:Gene: STY1180

Query Match 12.7%; Score 98.5; DB 2; Length 151;

Best Local Similarity 28.8%; Pred. No. 0.21;

Matches 34; Conservative 17; Mismatches 48; Indels 19; Gaps 5;

QY 51 GSANAALALQSDARKSE---TTITQSGYNGADVQ-GADNST-----IELTQ 94

Db 14 GAGCIATATNYDLARSEYNAFVNELSKSFNQAAIIGVGTDSARVRQEGSKLLSVISQ 73

QY 95 NGRFNATIDQNAKND-OLTRVVTTHMAHANTASDSSVMVQVGFNNATANY 151

Db 74 EGGNNRAKVDQ--AGNYPAYIEQTGNANDASISQAYGNSAAIIQKSGNKNATITQY 129

RESULT 10

S70787

curlin nucleator protein csb precursor - Escherichia coli (strain K-12)

N:Alternate names: csbB protein; curlin nucleation component; minor curlin protein
C:Species: Escherichia coli
C>Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 01-Mar-2002
C:Accession: S70787; F64846
R:Hammar, M.; Arngvist, A.; Bian, Z.; Olsen, A.; Normark, S.
Mol. Microbiol. 18, 661-670, 1995
A:Title: Expression of two csb operons is required for production of fibronectin- and C

A:Reference number: S70783; MUID:96414468; PMID:8817489

A:Accession: S70787

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-151 <HAM>

A:Cross-references: EMBL:X90754; NID:G1147558; PID:CAA62281.1; PID:G1147563

A:Experimental source: strain K12, substrain W310

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Ferna, N.T.; Burland, V.; Riley, M.; Co

.A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: F64846

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-151 <BLAT>

A:Cross-references: GB:AE000205; GB:U00096; NID:G1787265; PID:AA474125.1; PID:G1787278;

A:Experimental source: strain K-12, substrain MG1655

C:Genetics:

A:Gene: csbB

A:Map position: 23.15

C:Function:

A:Description: minor component of wild-type curli; interaction between CsgA and CsgB tri

A:Note: curli are thin, coiled fibers expressed on the surface of Escherichia coli that

and H-kininogen; in the absence of CsgA, CsgB can self-assemble into polymers

F:1-21/Domain: signal sequence #status predicted <SIG>

F:22-151/Product: minor curlin chain #status predicted <MAT>

Query Match 12.5%; Score 96.5; DB 2; Length 151;

Best Local Similarity 25.6%; Pred. No. 0.31;

Matches 30; Conservative 24; Mismatches 46; Indels 17; Gaps 4;

QY 51 GSANAALALQSDARKSE---TTITQSGYNGADVQGDADNSTIELTQGFNNATIDQ 105

Db 14 GAGCIATAAGYDLARSEYNAFVNELSKSFNQAAIIGVGTNSAQIRGGSKLLAVVQAQ 73

QY 106 WNAKYNQDLVTRVVTTHMAHANO--TASDSSV-----MVRQVGFNNATANY 151

Db 74 EGSSNRKAK-IDQTGDYLAVIDQAGSANDASISQAYGNTAMTIQKSGNKNATITQY 129

RESULT 11

C90806

minor curlin subunit precursor CsgB [imported] - Escherichia coli (strain O157:H7, subst

C:Species: Escherichia coli

C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001

C:Accession: C90806

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.

Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen

A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: C90806

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-151 <HAY>

A:Cross-references: GB:BA000007; PID:BA34842.1; PID:G13360879; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RIMD 0509952

C:Genetics:

A:Gene: ECs1419

Query Match 12.5%; Score 96.5; DB 2; Length 151;

Best Local Similarity 25.6%; Pred. No. 0.31;

Matches 30; Conservative 24; Mismatches 46; Indels 17; Gaps 4;

```
QY 51 GSANALALQSDARKSE-----TTTQSGYNGADVGGADNSTIELTQNGFRNNATIDQ 105
DB 14 GAGPFAAAGVDLANSEYNFVNELSKSFNQAAIIQAGTNNQAQLRQGGSKLLAVVAQ 73
QY 106 WNAKNYDQLVTRVVTHEMAHANQ--TASDSSV-----MVRQVGFNNATANQY 151
DB 74 EGSSNRK-IDQTGDYNLAYIDQAGSANDASISQAGYNTAMIIQKSGNKNANTIQY 129

RESULT 12
G85665
Curlin minor chain precursor, CsgA homolog [imported] - Escherichia coli (strain O157:H7)
C:Species: Escherichia coli
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: G85665
R:Ripera, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
  Miller, L.; Grobbeck, B.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
  Nature 409, 529-533, 2001
A>Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: G85665
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-151 <STO>
A:Cross-references: GB:AE005174; NID:G12514573; PIDN:AAG55787.1; GSPDB:GN00145; UWGP:Z16
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: csgB

Query Match 12.5%; Score 96.5; DB 2; Length 151;
Best Local Similarity 25.6%; Pred. No. 0.31;
Matches 30; Conservative 24; Mismatches 46; Indels 17; Gaps 4;

QY 51 GSANALALQSDARKSE-----TTTQSGYNGADVGGADNSTIELTQNGFRNNATIDQ 105
DB 14 GAGPFAAAGVDLANSEYNFVNELSKSFNQAAIIQAGTNNQAQLRQGGSKLLAVVAQ 73
QY 106 WNAKNYDQLVTRVVTHEMAHANQ--TASDSSV-----MVRQVGFNNATANQY 151
DB 74 EGSSNRK-IDQTGDYNLAYIDQAGSANDASISQAGYNTAMIIQKSGNKNANTIQY 129

RESULT 13
D70604
probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: D70604
R:Coale, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
  Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
  Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
  Nature 393, 537-544, 1998
A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: D70604
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-552 <COL>
A:Cross-references: GB:Z92774; GB:AL123456; NID:g3261729; PIDN:CAB07133.1; PID:g1877289
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: PPE
C:Superfamily: Phaseolus glycine-rich cell wall protein 1.8

Query Match 12.1%; Score 93.5; DB 2; Length 552;
Best Local Similarity 27.9%; Pred. No. 2.4;
Matches 41; Conservative 11; Mismatches 58; Indels 37; Gaps 7;

QY 15 SGSAALAGV-----VPQWGGGNGHNGGNSGPD-----STLSIYQYGSANAALALQSDA 63
DB 353 SGSGNIGVNTGANTLVP---GDLNLLGVGNSGANIGFAGNAGVLTGFGNASTLTGLG 409
```

```
QY 64 RKSETTITQSGYNGADVGGADNSTIELTQNGFRNNATIDQNAKYDQ---LVTRVVT 120
DB 410 NAGE---LMTFGNAGFVNTGFDNSGNVNTGNSGNINTGSMNAGNVTGFGIIT---- 462
QY 121 HEMAHANQATASDSSVMVRQVQFGNNAT 147
DB 463 -----DSG--LTNSGFGNTGT 476

RESULT 14
AD1539
probable sugar ABC transporter, periplasmic sugar-binding protein homolog lin0852 [impo
C:Species: Listeria innocua
C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C:Accession: AD1539
R:Glaser, P.; Franguel, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecke
  ; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, C.; Entian, K.D.; Fsihi, H.
  D.; Jones, L.M.; Karst, U.
  Science 294, 849-852, 2001
A:Authors: Krefte, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitouriam, A.; M
  ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland
  A>Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AD1539
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-440 <GLA>
A:Cross-references: GB:AL592022; PIDN:CAC96084.1; PID:g16413303; GSPDB:GN00178
A:Experimental source: strain Clp11262
C:Genetics:
A:Gene: lin0852

Query Match 11.9%; Score 92.5; DB 2; Length 440;
Best Local Similarity 23.8%; Pred. No. 2.3;
Matches 38; Conservative 23; Mismatches 59; Indels 43; Gaps 5;

QY 1 MKLLKVA-AFAAIVSGSALAGVVPQWGGGNGHNGGNN----- 37
DB 1 MKRKIAIAIALSVAVAGSLLTAC---GGGKSGDNGTKVTFWAAPNPTQVKYWDEM 56
QY 38 -----SSGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGADVGGADNSTIEL 92
DB 57 AKAYEKENPDVTVEVSQMKESPSEATIOSAIAKTAAPTSENINSPAAQLADSKAI-V 115
QY 93 TONGFRN-----NATIDQNAKYDQLVTRVVTHEMAHA 126
DB 116 PLNDVRGLDDVVKRWKXSETMDSKFKSDGNQYVLPVYSNPILFA 159

RESULT 15
A56038
DNA-binding protein ovo - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C>Date: 01-Dec-1995 #sequence_revision 01-Dec-1995 #text_change 21-Jul-2000
C:Accession: A56038
R:Garfinkel, M.D.; Wang, J.; Liang, Y.; Mahowald, A.P.
  Mol. Cell. Biol. 14, 6809-6818, 1994
A>Title: Multiple products from the shavenbaby-ovo gene region of Drosophila melanogast
A:Reference number: A56038; MUID:95021209; PMID:7935398
A:Accession: A56038
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1028 <GAR>
A:Cross-references: GB:U11383; NID:g520526; PIDN:AAB60216.1; PID:g520527
C:Genetics:
A:Gene: ovo
A:Cross-references: FlyBase:FBgn0003028

Query Match 11.9%; Score 92.5; DB 2; Length 1028;
Best Local Similarity 26.7%; Pred. No. 5.9;
Matches 40; Conservative 14; Mismatches 61; Indels 35; Gaps 6;

QY 3 LLKVAFAAIVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQSD 62
```

```

Db      59  LONAAAAAYIMSAGSG-----GGCTGNGGGGSGPGGPGPSANSGGGGGG----- 104
Qy      63  ARKSETTITQSGYNGADYQGADNSTIELTQNGFRNNATIQQWNAKNDYDQIVTRVVTHE 122
Db      105  -----GGGYINCXGVG-GPNNS---LDGNNLLNFPASVSNYNESN-----SKFHNEH 147
Qy      123  MAHA-----NOTASDSSVVRQVFGNNATA 148
Db      148  HHQHNNNNNNNGGQTSMMGHPFYGGNPSA 177

```

Search completed: August 2, 2004, 14:56:24
Job time : 9.4 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:36:12 ; Search time 5.3 seconds

(without alignments)
1483.508 Million cell updates/sec

Title: US-09-543-407-24

Perfect score: 775

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSVNVVRQVGFNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	690	89.0	151	1 CSGA_SALTY	P55225 salmonella
2	533	68.8	151	1 CSGA_ECOLI	P28307 escherichia
3	511.5	66.0	152	1 CSGA_ECO57	Q93u24 escherichia
4	98.5	12.7	151	1 CSGB_SALTY	Q87m3 salmonella
5	98.5	12.7	151	1 CSGB_SALTY	P55226 salmonella
6	96.5	12.5	151	1 CSGB_ECOLI	P39828 escherichia
7	92.5	11.9	1028	1 CVO_DROME	P51521 drosophila
8	90.5	11.7	1327	1 TNK1_HUMAN	Q95271 homo sapien
9	90	11.6	347	1 MSA2_PLAF2	Q03646 plasmodium
10	89	11.5	262	1 VG38_BPT2	P07875 bacterioph
11	87	11.2	590	1 GP63_LEIDO	P23223 leishmania
12	87	11.2	599	1 GP63_LEICH	P15706 leishmania
13	87	11.2	602	1 GP63_LEIMA	P08148 leishmania
14	87	11.2	646	1 GP63_LEIME	P43150 leishmania
15	87	11.2	1656	1 OMPB_RICUA	P06653 x outer mem
16	85	11.0	401	1 YK03_CABEL	P34291 caenorhabdi
17	82	10.6	678	1 YF48_MVCTU	Q10778 mycobacteri
18	82	10.6	1093	1 PER_BROWI	Q03297 drosophila
19	82	10.6	1115	1 TB22_CHLRE	Q8xrp3 chlamydonon
20	81.5	10.5	306	1 HMXD_DROPS	P20822 drosophila
21	81.5	10.5	311	1 HMXD_PSESP	P07662 pseudomonas
22	81.5	10.5	392	1 HME1_HUMAN	Q05925 homo sapien
23	81	10.5	165	1 GRP1_ORYSA	Q25074 oryza sativ
24	80.5	10.4	548	1 CEAK_ECOLI	Q47502 escherichia
25	80.5	10.4	720	1 G7AK_BREDI	Q915d6 brevedimonon
26	80.5	10.4	894	1 ILF3_HUMAN	Q12906 h interleuk
27	80	10.3	427	1 CFIA_DROME	P16241 drosophila
28	80	10.3	491	1 YK99_MYCTU	Q10707 mycobacteri
29	79.5	10.3	172	1 CH18_DROME	P07184 drosophila
30	79.5	10.3	760	1 YBIL_ECOLI	P75780 escherichia
31	79.5	10.3	1034	1 ICEN_PANAN	Q47879 pantoea ana
32	79.5	10.3	1258	1 ICEN_ERWHE	P16239 erwina her
33	79.5	10.3	1322	1 ICFA_PANAN	P20469 pantoea ana

RESULT 1

CSGA_SALTY

ID	CSGA_SALTY	STANDARD;	PRT;	151 AA.
AC	P55225;			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Major curlin subunit precursor (Fimbrin SBE17).			
GN	CSGA OR AGFA OR STM1144 OR STY1181 OR T1776.			
OS	Salmonella typhimurium,			
OS	Salmonella typhi, and			
OS	Salmonella enteritidis.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OC	Enterobacteriaceae; Salmonella.			
OX	NCBI_TaxID=602, 601, 592;			
OX	[1]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=S.typhimurium; STRAIN=SR-11;			
RC	MEDLINE=98117058; PubMed=9457880;			
RA	Romling U., Bian Z., Hammar M., Sierralta W.D., Normark S.;			
RA	"Curli fibers are highly conserved between Salmonella typhimurium and			
RT	Escherichia coli with respect to operon structure and regulation.";			
RL	J. Bacteriol. 180:722-731(1998).			
RL	[2]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=S.typhimurium; STRAIN=L72 / SGSC1412 / ATCC 700720;			
RC	MEDLINE=21534948; PubMed=11677609;			
RA	McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,			
RA	Courtney L., Porwollik S., Ali J., Dante M., Du P., Hou S., Layman D.,			
RA	Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,			
RA	Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,			
RA	Waterston R., Wilson R.K.;			
RT	"Complete genome sequence of Salmonella enterica serovar Typhimurium			
RT	LT2.";			
RL	Nature 413:852-856(2001).			
RL	[3]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=S.typhi; STRAIN=CT18;			
RC	MEDLINE=21534947; PubMed=11677608;			
RA	Parhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,			
RA	Churcher C., Mungall K.L., Bentley S.D., Holden M.R.T., Sebahia M.,			
RA	Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,			
RA	Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,			
RA	Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jageis K.,			
RA	Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,			
RA	Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,			
RA	Whitehead S., Barrall B.G.;			
RT	"Complete genome sequence of a multiple drug resistant Salmonella			
RT	enterica serovar Typhi CT18.";			
RL	Nature 413:848-852(2001).			
RL	[4]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=S.typhi; STRAIN=Ty2 / ATCC 700931;			
RC	MEDLINE=22531367; PubMed=12644504;			
RA	Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,			
RA	Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;			

34	79	10.2	878	1	FIMD_ECOLI
35	79	10.2	933	1	NPA3_HUMAN
36	78.5	10.1	718	1	FLGE_HELPJ
37	78.5	10.1	718	1	FLGE_HELPY
38	78	10.1	1087	1	SGG_DROME
39	77	9.9	796	1	MEN_DROME
40	77	9.9	1146	1	YHC3_YEAST
41	77	9.9	1690	1	KFLA_HUMAN
42	76.5	9.9	348	1	YZ17_AQUAE
43	76.5	9.9	979	1	RFX1_HUMAN
44	76.5	9.9	1196	1	ICEV_PSEEX
45	76.5	9.9	1210	1	ICEN_PSEFL

ALIGNMENTS

"Comparative genomics of *Salmonella enterica* serovar Typhi strains Ty2 and CT18";
 J. Bacteriol. 185:2330-2337(2003).
 [5]
 RN SEQUENCE FROM N.A.
 RP SPECIES-S. enteritidis; STRAIN=27655-3B;
 RC MEDLINE=96146512; PubMed=8550497;
 RA Collinson S.K., Clouthier S.C., Doran J.L., Baner P.A., Kay W.W.;
 RT "Salmonella enteritidis agfBAC operon encoding thin, aggregative
 fimbriae";
 J. Clin. Microbiol. 178:662-667(1996).
 [6]
 RN SEQUENCE OF 21-151 FROM N.A.
 RP SPECIES-S. enteritidis; STRAIN=27655-3B;
 RC MEDLINE=94013373; PubMed=8104955;
 RA Doran J.L., Collinson S.K., Burian J., Santos G., Todd E.C.D.,
 Munro C.K., Kay W.W., Baner P.A., Peterkin P.I., Kay W.W.;
 RT "DNA-based diagnostic tests for *Salmonella enteritidis* targeting agfA,
 the structural gene for thin, aggregative fimbriae";
 J. Clin. Microbiol. 31:2263-2273(1993).
 [7]
 RN SEQUENCE OF 21-33.
 RP SPECIES-S. enteritidis; STRAIN=27655-3B;
 RC MEDLINE=9310586; PubMed=1677357;
 RA Collinson S.K., Emedy L., Mueller K.-M., Trust T.J., Kay W.W.;
 RT "Purification and characterization of thin, aggregative fimbriae from
Salmonella enteritidis";
 J. Bacteriol. 173:4773-4781(1991).
 CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 FIBRONECTIN.
 CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 DR EMBL; AJ002301; CAA05317.1; -;
 DR EMBL; AF008749; AAL20074.1; -;
 DR EMBL; AL627469; CAD08268.1; -;
 DR EMBL; AS016840; AAC09399.1; -;
 DR EMBL; U43280; AAC43599.1; -;
 DR FIR; JC6039; JC6039.
 DR StyGene; SG10608; csGA.
 KW Fimbria; Signal; Complete proteome.
 FT SIGNAL 1 20
 FT CHAIN 21 151 MAJOR CURLIN SUBUNIT.
 FT CONFLICT 134 151 SVMVQVGFNNATANYQ -> DSVTQVAS (IN
 REF. 6).
 SQ SEQUENCE 151 AA; 15305 MW; B7DAC0D16B621359 CRC64;
 Query Match 89.0%; Score 690; DB 1; Length 151;
 Best Local Similarity 91.4%; Pred. No. 1.4e-51;
 Matches 138; Conservative 2; Mismatches 11; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGNGNGSGPDSITLSIYQYGSANAALALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGNGNGSGPDSITLSIYQYGSANAALALQ 60
 QY 61 SDARKSETTITQSGYNGADVGGADNSTIETQNGFRNNATIDOWNAKNYDOLVTRVVT 120
 DB 61 SDARKSETTITQSGYNGADVGGADNSTIETQNGFRNNATIDOWNAKNSDITVQYGG 120
 QY 121 HEMAHANQASDSSVMVQVGFNNATANYQ 151
 DB 121 NNAALVNQASDSSVMVQVGFNNATANYQ 151

RESULT 2
 CSGA_ECOLI
 ID _CSGA_ECOLI STANDARD; PRT; 151 AA.
 AC P28307;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Major curlin subunit precursor.
 GN CSGA OR B1042.
 OS *Escherichia coli*.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; *Escherichia*.
 OX NCBI_TaxID=562;
 [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=K12 / W3110;
 RC MEDLINE=93211294; PubMed=8459772;
 RX Olsen A., Arqvist A.;
 RA "The RpoS sigma factor relieves H-NS-mediated transcriptional
 repression of csGA, the subunit gene of fibronectin-binding curli in
Escherichia coli";
 Mol. Microbiol. 7:523-536(1993).
 [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN=K12 / MC4100;
 RC MEDLINE=96414469; PubMed=8817489;
 RX Hammar M., Arqvist A., Bian Z., Olsen A., Normark S.;
 RA "Expression of two csG operons is required for production of
 fibronectin- and Congo red-binding curli polymers in *Escherichia coli*
 K-12";
 Mol. Microbiol. 18:661-670(1995).
 [3]
 RN SEQUENCE FROM N.A.
 RP STRAIN=K12 / MG1655;
 RC MEDLINE=97426617; PubMed=92798503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 Mau B., Shao Y.;
 RA "The complete genome sequence of *Escherichia coli* K-12";
 Science 277:1453-1474(1997).
 [4]
 RN SEQUENCE FROM N.A.
 RP STRAIN=K12;
 RC MEDLINE=97061202; PubMed=8905232;
 RX Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
 Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
 Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
 Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
 Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 Yano M., Horiuchi T.;
 RA "A 718-kb DNA sequence of the *Escherichia coli* K-12 genome
 corresponding to the 12.7-28.0 min region on the linkage map";
 DNA Res. 3:137-155(1996).
 [5]
 RN SEQUENCE OF 21-40.
 RP STRAIN=K12 / YMEL;
 RC MEDLINE=93023873; PubMed=1357528;
 RX Arqvist A., Olsen A., Pfeifer J., Russell D.G., Normark S.;
 RT "The Crl protein activates cryptic genes for curli formation and
 fibronectin binding in *Escherichia coli* HB101";
 Mol. Microbiol. 6:2443-2452(1992).
 [6]
 RN SEQUENCE OF 21-31.
 RP MEDLINE=91310586; PubMed=1677357;
 RX Collinson S.K., Emedy L., Trust T.J., Kay W.W.;
 RA "Purification and characterization of thin, aggregative fimbriae from
Salmonella enteritidis";
 J. Bacteriol. 173:4773-4781(1991).
 CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 FIBRONECTIN.
 CC

```

CC      -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      DR EMBL; L04979; AAR23616.1; -
CC      DR EMBL; X90754; CAA62282.1; -
CC      DR EMBL; AE000205; AAC74126.1; -
CC      DR EMBL; D90741; BAA35832.1; -
CC      DR EMBL; D90742; BAA35840.1; -
CC      DR PIR; S70788; S70788.
CC      DR EcoGene; EG11489; csga.
CC      KW Fimbria; Signal; Complete proteome.
CC      FT SIGNAL 1 20
CC      FT CHAIN 21 151 MAJOR CURLIN SUBUNIT.
CC      FT CONFLICT 7 A -> E (IN REF. 1).
CC      SQ SEQUENCE 151 AA; 15049 MW; C003470D208D395F CRC64;

Query Match 68.8%; Score 533; DB 1; Length 151;
Best Local Similarity 70.2%; Pred. No. 2.4e-38;
Matches 106; Conservative 18; Mismatches 27; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQVGGGNGHNGSGSPDSTLSIYYGGSANAALALQ 60
DQ 1 MKLLKVAAPAAIVVSGSALAGVVPQVGGGNGHNGSGSPDSTLSIYYGGSANAALALQ 60
QY 61 SDARKSETITQSGYNGADYVGGADNSITELTQNGFNNTATIDQWNAKYDQVTRVV 120
DQ 61 TDARNSDLTITQHGNGGADYVGGSDSDITLQKFGNSATLDQWNGKSNMTVKQFG 120
QY 121 HEMAHANTASDSSVMVRQVGFNNATANQY 151
DQ 121 GNGAAVDQTASNSVTVTVQVGFNNATANQY 151

RESULT 3
CSGA_ECO57
ID CSGA_ECO57 STANDARD; PRT; 152 AA.
AC Q93U24;
DC 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Major curlin subunit precursor.
GN CSGA OR Z1676 OR EC81420.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / ATCC 43895;
RX MEDLINE=21218556; PubMed=11319125;
RA Ulrich G.A., Keen J.E., Elder R.O.;
RT "Mutations in the csdG promoter associated with variations in curli
RT expression in certain strains of Escherichia coli O157:H7."
RL Appl. Environ. Microbiol. 67:2367-2370(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11208551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamocis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
RT Nature 409:529-533(2001).

```

```

[3]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12."
RL DNA Res. 8:11-22(2001).
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN.
CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      DR EMBL; AF275733; AAK53212.1; -
CC      DR EMBL; AE005315; AAG55788.1; -
CC      DR EMBL; AP002554; BAB34843.1; -
CC      DR PIR; D90806; D90806.
CC      DR PIR; H85665; H85665.
CC      KW Fimbria; Signal; Complete proteome.
CC      FT SIGNAL 1 20 BY SIMILARITY.
CC      FT CHAIN 21 152 MAJOR CURLIN SUBUNIT.
CC      SQ SEQUENCE 152 AA; 15099 MW; EE2D294DDE91243 CRC64;

Query Match 66.0%; Score 511.5; DB 1; Length 152;
Best Local Similarity 68.4%; Pred. No. 1.5e-36;
Matches 104; Conservative 19; Mismatches 28; Indels 1; Gaps 1;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQVGGGNGHNGSGSPDSTLSIYYGGSANAALAL 59
DQ 1 MKLLKVAAPAAIVVSGSALAGVVPQVGGGNGHNGSGSPDSTLSIYYGGSANAALAL 60
QY 60 OSDARKSETITQSGYNGADYVGGADNSITELTQNGFNNTATIDQWNAKYDQVTRVV 119
DQ 61 QADARNSDLTITQHGNGGADYVGGSDSDITLQKFGNSATLDQWNGKSNMTVKQFG 120
QY 120 THEMAHANTASDSSVMVRQVGFNNATANQY 151
DQ 121 GNGAAVDQTASNSVTVTVQVGFNNATANQY 152

RESULT 4
CSGB_SALTI
ID CSGB_SALTI STANDARD; PRT; 151 AA.
AC Q9Z7M3;
DC 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Minor curlin subunit precursor.
GN CSGB OR STY1180 OR T1777.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Conerton P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,

```

RA Feltwell T., Hamlin N., Haque A., Hien T.T., Helroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrell B.G.,
RT "Complete genome sequence of a multiple drug resistant Salmonella
enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Aji J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
LT2.";
RL Nature 413:852-856(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=S.enteritidis; STRAIN=27655-3B;
RX MEDLINE=96146512; PubMed=8550497;
RA Collinson S.K., Clouthier S.C., Doran J.L., Baner P.A., Kay W.W.;
RT "Salmonella enteritidis agfBAC operon encoding thin, aggregative
fimbriae.";
RL J. Bacteriol. 178:662-667(1996).
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
CURLIN MONOMERS.
CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; AJ002301; CAA05316.1; -
CC EMBL; AE008749; AAL20073.1; -
CC EMBL; U43280; AAC43598.1; -
CC PIR; JC6040; JC6040.
CC StyGene; SG10609; csGB.
CC Fimbrin; Signal; Complete proteome.
CC SIGNAL 21 POTENTIAL.
CC CHAIN 22 151 MINOR CURLIN SUBUNIT.
CC SEQUENCE 151 AA; 16182 MW; 60FC5430E6DD361D CRC64;
Query Match 12.7%; Score 98.5; DB 1; Length 151;
Best Local Similarity 28.8%; Pred. No. 0.099;
Matches 34; Conservative 17; Mismatches 48; Indels 19; Gaps 5;
QY 51 GSANAALALQSDARKSE-----TTITQSGYNGADVQ-GADNST-----IELTQ 94
Db 14 GAGCIATATNYDLARSEYFANVELSKSFNQAAIIGQVGTDSARVRQEGSKLSVISQ 73
QY 95 NGRFNATIDOWNAKNYD-QLVTRVVTHEMAHANOTASDSSVMVROVGFNNATANOY 151
Db 74 EGGNNRAKVDQ--AGNYNFAYIEQTGNANDASISQAYGNSAAIIKQSGKNANTIOY 129
RESULT 6
CSGB_ECOLI STANDARD; PRT; 151 AA.
ID CSGB_ECOLI
AC P39828;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Minor curlin subunit precursor.
GN CSGB OS B1041 OR Z1675 OR ECS1419.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]

RA Feltwell T., Hamlin N., Haque A., Hien T.T., Helroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrell B.G.,
RT "Complete genome sequence of a multiple drug resistant Salmonella
enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=LT2 / ATCC 700720;
RX MEDLINE=22531367; PubMed=126454504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodoyanni V., Schwartz D.C., Blattner F.R.;
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
and CT18.";
RL J. Bacteriol. 185:2330-2337(2003).
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
CURLIN MONOMERS.
CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; AL627269; CAD08267.1; -
CC EMBL; AE016840; AAC09400.1; -
CC Fimbrin; Signal; Complete proteome.
CC SIGNAL 21 POTENTIAL.
CC CHAIN 22 151 MINOR CURLIN SUBUNIT.
CC SEQUENCE 151 AA; 16254 MW; 161C54326E573495 CRC64;
Query Match 12.7%; Score 98.5; DB 1; Length 151;
Best Local Similarity 28.8%; Pred. No. 0.099;
Matches 34; Conservative 17; Mismatches 48; Indels 19; Gaps 5;
QY 51 GSANAALALQSDARKSE-----TTITQSGYNGADVQ-GADNST-----IELTQ 94
Db 14 GAGCIATATNYDLARSEYFANVELSKSFNQAAIIGQVGTDSARVRQEGSKLSVISQ 73
QY 95 NGRFNATIDOWNAKNYD-QLVTRVVTHEMAHANOTASDSSVMVROVGFNNATANOY 151
Db 74 EGGNNRAKVDQ--AGNYNFAYIEQTGNANDASISQAYGNSAAIIKQSGKNANTIOY 129
RESULT 5
CSGB_SALTY STANDARD; PRT; 151 AA.
ID CSGB_SALTY
AC P55226;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Minor curlin subunit precursor (Fimbrin SEF17 minor subunit).
GN CSGB OR AGFB OR STM143.
OS Salmonella typhimurium, and
OS Salmonella enteritidis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602, 592;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=SR-11;
RX MEDLINE=98117058; PubMed=9457880;
RA Romling U., Bian Z., Hammar M., Sierralta W.D., Normark S.;
RT "Curli fibers are highly conserved between Salmonella typhimurium and
Escherichia coli with respect to operon structure and regulation.";
RL J. Bacteriol. 180:722-731(1998).

RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MC4100;
RX MEDLINE=96414468; PubMed=8817489;
RA Hammar M., Arngvist A., Bian Z., Olsen A., Normark S.;
RT "Expression of two csg operons is required for production of
RT fibronectin- and Congo red-binding curli polymers in *Escherichia coli*
RT K-12";
RL Mol. Microbiol. 18:661-670 (1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of *Escherichia coli* K-12";
RL Science 277:1453-1474 (1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / ATCC 700927;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RT "A 718-Kb DNA sequence of the *Escherichia coli* K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map";
RL DNA Res. 3:137-155 (1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimallanta E.T., Potamouzis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7";
RL Nature 409:529-533 (2001).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic *Escherichia coli*
RT O157:H7 and genomic comparison with a laboratory strain K-12";
RL DNA Res. 8:11-22 (2001).
RN [6]
RP SEQUENCE OF 1-21 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=95157246; PubMed=7854117;
RA Arngvist A., Olsen A., Normark S.;
RT "Sigma S-dependent growth-phase induction of the *csgBA* promoter in
RT *Escherichia coli* can be achieved in vivo by sigma 70 in the absence
RT of the nucleoid-associated protein H-NS";
RL Mol. Microbiol. 13:1021-1032 (1994).
RN [7]
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
CC CURLIN MONOMERS.
CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X90754; CAA62281.1; -
CC EMBL; AE000205; AAC74125.1; -
CC EMBL; D90741; BAA35831.1; -
CC EMBL; AE005315; AAG55787.1; -
CC EMBL; AP002554; BAB34842.1; -
CC PIR; C90806; C90806;
CC PIR; G85665; G85665;
CC PIR; S70787; S70787;
CC EcoGene; EG12621; csgB.
CC FimBria; Signal; complete proteome.
CC SIGNAL 1 21
CC FT CHAIN 22 151 MINOR CURLIN SUBUNIT.
CC SEQUENCE 151 AA; 15682 MW; B18D266B964014B8 CRC64;
CC
CC Query Match 12.5%; Score 96.5; DB 1; Length 151;
CC Best Local Similarity 25.6%; Pred. No. 0.15;
CC Matches 30; Conservative 24; Mismatches 46; Indels 17; Gaps 4;
CC
CC QY 51 GSANAAALQSDARKS-----TTITQSGYGVGADVGQADNSTIETLQNGFRNNATIDQ 105
CC Db 14 GAPGIAAAGYDLANSEYFVAVNELSKSFNQAAIICQAGTNSAQLRQSGSKLLAVVAQ 73
CC QY 106 WNAKNYDQVTRVVTHEMAHANQ--TASDSV-----MVRQVGFNNATANQY 151
CC Db 74 EGSSNRK- IDQTDYNLAVIDQAGSANDASISQGYNTAMTIQSGNKNANTQY 129
CC
CC RESULT 7
CC OVO DROME
CC ID OVO DROME STANDARD; PRT; 1028 AA.
CC AC P51521; Q9XZU4;
CC DT 01-OCT-1996 (Rel. 34, Created)
CC DT 01-OCT-1996 (Rel. 34, Last sequence update)
CC DT 10-OCT-2003 (Rel. 42, Last annotation update)
CC DE OVO protein (Shaven baby protein).
CC GN OVO OR SVB.
CC OS *Drosophila melanogaster* (Fruit fly).
CC SC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
CC OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
CC OC Ephydroidea; Drosophilidae; Drosophila.
CC OX NCBI_TaxID=7227;
CC [1]
CC RP SEQUENCE FROM N.A.
CC RX TISSUE=Ovary;
CC RC MEDLINE=95021209; PubMed=7935398;
CC RA Garfinkel M.D., Wang J., Liang Y., Mahowald A.P.;
RT "Multiple products from the shavenbaby-ovc gene region of *Drosophila*
RT melanogaster: relationship to genetic complexity";
RL Mol. Cell. Biol. 14:6809-6818 (1994).
RN [2]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN=Oregon-R;
CC RX MEDLINE=91293102; PubMed=1712294;
CC RA Mevel-Ninio M.T.M., Terracol R., Kafatos F.C.;
RT "The ovo gene of *Drosophila* encodes a zinc finger protein required
RT for female germ line development";
RL EMBO J. 10:2259-2266 (1991).
RN [3]
CC -!- FUNCTION: REQUIRED FOR SURVIVAL AND DIFFERENTIATION OF FEMALE GERM
CC LINE CELLS. PLAYS A ROLE IN GERM LINE SEX DETERMINATION.
CC -!- SUBCELLULAR LOCATION: Nuclear (potential).
CC -!- DEVELOPMENTAL STAGE: FIRST APPEARS IN THE GERMARIIUM AND
CC ACCUMULATES IN NURSE CELLS DURING OOGENESIS. STORED IN THE EGG,
CC BUT IS RAPIDLY LOST IN THE EMBRYOS EXCEPT FOR ITS CONTINUED
CC PRESENCE IN THE GERM LINE PRECURSOR POLE CELLS.
CC -!- SIMILARITY: Contains 4 C2H2-type zinc fingers.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; U11383; AAB60216.1; -;
DR EMBL; X59772; CAB36921.1; ALT_SEQ.
DR PIR; A56038; A56038.
DR HSP; P07248; 2ADR.
DR TRANSFAC; T00669; -;
DR FlyBase; FBgn0003028; ovo.
DR InterPro; IPR007087; znf_C2H2.
DR Pfam; PF000096; znf_C2H2; 3.
DR SMART; SM00355; ZNF_C2H2; 4.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 3.
KW Zinc-finger; Metal-binding; DNA-binding; Repeat; Nuclear protein;
KW Transcription regulation.
FT DOMAIN 62 66 POLY-ALA.
FT DOMAIN 72 77 POLY-GLY.
FT DOMAIN 80 85 POLY-GLY.
FT DOMAIN 98 108 POLY-GLY.
FT DOMAIN 144 152 POLY-HIS.
FT DOMAIN 153 159 POLY-ASN.
FT DOMAIN 336 339 POLY-GLN.
FT DOMAIN 347 353 POLY-GLN.
FT DOMAIN 357 361 POLY-GLN.
FT DOMAIN 410 414 POLY-GLN.
FT DOMAIN 418 422 POLY-GLN.
FT DOMAIN 426 432 POLY-GLN.
FT DOMAIN 445 453 POLY-GLN.
FT DOMAIN 456 459 POLY-GLN.
FT DOMAIN 466 474 POLY-GLN.
FT DOMAIN 497 517 POLY-ALA.
FT DOMAIN 524 529 POLY-SER.
FT DOMAIN 549 558 POLY-ALA.
FT DOMAIN 639 651 POLY-ALA.
FT DOMAIN 717 725 POLY-ALA.
FT DOMAIN 797 802 POLY-GLN.
FT DOMAIN 820 832 POLY-GLN.
FT DOMAIN 826 832 POLY-GLN.
FT ZN_FING 874 896 C2H2-TYPE 1.
FT ZN_FING 902 924 C2H2-TYPE 2.
FT ZN_FING 930 953 C2H2-TYPE 3.
FT ZN_FING 969 992 C2H2-TYPE 4.
FT ZN_FING 647 647 A -> R (IN REF. 2).
SQ SEQUENCE 1028 AA; 110620 MW; D7068B2BC0F6F77 CRC64;

Query Match 11.9%; Score 92.5; DB 1; Length 1028;
Best Local Similarity 26.7%; Pred. No. 2.8;
Matches 40; Conservative 14; Mismatches 61; Indels 35; Gaps 6;
QY 3 LKVAFAATVSGSALAGVVPWGGGNGHNGGSSGPDSTLSIYQYGSANAALALQSD 62
Db 59 LQNAARAAIYMSAGS-----GGCTGNGGGGASGPGPSANSNGGGGG----- 104
QY 63 ARKSETITQSGNGADVCGGADNSTIELTQNGFRNNTIDQWAKNYDQLVTRVTHE 122
Db 105 -----GGGYINGCGVG-GPNNS---LDGNLLNFASVSNYESN-----SKPHNH 147
QY 123 MAHA---NQTSADSSVMVRQVFGNNATA 148
Db 148 HHQHNNNNNNNGQTSMGHPYGGNPSA 177

RESULT 8
ID_TNKL_HUMAN STANDARD; PRT; 1327 AA.
AC O95271; O95272;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE DE Tankyrase 1 (EC 2.4.2.30) (TANK1) (Tankyrase I) (TNKS-1) (TRF1-
DE interacting ankyrin-related ADP-ribose polymerase).
CN CN TNKS OR TNKS1 OR TIN1 OR TINF1 OR PARP1.
OS OS Homo sapiens (Human).
OC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX OX NCBI_TaxID=9606;
RN RN SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RP RP TISSUE=Testis;
RC RC MEDLINE=99454782; PubMed=10523501;
RX RX Smith S., Giriat I., Schmitt A., de Lange T.;
RA RA "Tankyrase, a poly(ADP-ribose) polymerase at human telomeres.";
RT RT Science 282:1484-1487(1998).
RL RL [2]
RN RN SUBCELLULAR LOCATION.
RP RP MEDLINE=99454782; PubMed=10523501;
RX RX Smith S., de Lange T.;
RA RA "Cell cycle dependent localization of the telomeric PARP, tankyrase,
RT RT to nuclear pore complexes and centrosomes.";
RL RL J. Cell Sci. 112:3649-3656(1999).
RN RN [3]
RP RP FUNCTION AND PHOSPHORYLATION.
RX RX MEDLINE=20556282; PubMed=10988299;
RA RA Chi N.-W., Lodish H.F.;
RT RT "Tankyrase is a Golgi-associated mitogen-activated protein kinase
RL RL substrate that interacts with IRAP in GLUT4 vesicles.";
RN RN J. Biol. Chem. 275:38437-38444(2000).
RP RP [4]
RX RX FUNCTION AND MUTAGENESIS OF HIS-1184 AND GLU-1291.
RA RA MEDLINE=21602874; PubMed=11739745;
RT RT Cook B.D., Dynek J.N., Chang W., Shostak G., Smith S.;
RL RL "Role for the related poly(ADP-Ribose) polymerases tankyrase 1 and 2
RN RN at human telomeres.";
RP RP J. Mol. Biol. 22:332-342(2002).
RX RX -1- FUNCTION: May regulate vesicle trafficking and modulate the
RA RA subcellular distribution of SLC2A4/GLUT4-vesicles. Has PARP
RT RT activity and can modify TRF1, and thereby contribute to the
RL RL regulation of telomere length.
RN RN -1- CATALYTIC ACTIVITY: NAD(+) + {ADP-D-ribose}(N)-acceptor =
RX RX nicotinamide + {ADP-D-ribose}(N+1)-acceptor.
RA RA -1- SUBUNIT: Oligomerizes and associates with TNKS2. Interacts with
RT RT the cytoplasmic domain of LNPEP/Otase in SLC2A4/GLUT4-vesicles.
RL RL Binds to the N-terminus of telomeric TRF1 via the ANK repeats.
RN RN -1- SUBCELLULAR LOCATION: Cytoplasmic, associated with the Golgi and
RX RX with juxtanuclear SLC2A4/GLUT4-vesicles. A minor proportion is
RA RA also found at nuclear pore complexes and around the pericentriolar
RT RT matrix of mitotic centrosomes. During interphase, a small fraction
RL RL of TNKS is found in the nucleus, associated with TRF1.
RN RN -1- ALTERNATIVE PRODUCTS:
RX RX Event=Alternative splicing; Named isoforms=2;
RA RA Name=1;
RT RT IsoId=O95271-1; Sequence=Displayed;
RL RL Name=2;
RN RN IsoId=O95271-2; Sequence=VSP_004538, VSP_004539;
RX RX Note=No experimental confirmation available;
RA RA -1- TISSUE SPECIFICITY: Ubiquitous; highest levels in testis.
RT RT -1- PTM: Upon insulin-stimulation, phosphorylated on serine residues
RL RL by MAPK kinases.
RN RN -1- PTM: ADP-ribosylated (-auto).
RX RX -1- SIMILARITY: Belongs to the PARP family.
RA RA -1- SIMILARITY: Contains 15 ANK repeats.
RT RT -1- SIMILARITY: Contains 1 sterile alpha motif (SAM) domain.
RL RL -----
RN RN This SWISS-PROT entry is copyright. It is produced through a collaboration
RX RX between the Swiss Institute of Bioinformatics and the EMBL outstation -
RA RA the European Bioinformatics Institute. There are no restrictions on its
RT RT use by non-profit institutions as long as its content is in no way
RL RL modified and this statement is not removed. Usage by and for commercial
RN RN entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
RX RX or send an email to license@isb-sib.ch).
RA RA -----
RT RT EMBL; AF082556; AAC79841.1; -;
RL RL -----

RA Miller R.A., Reed S.G., Parsons M.;
 RT "Leishmania gp63 molecule implicated in cellular adhesion lacks an
 RL Arg-Gly-Asp sequence.";
 RL Mol. Biochem. Parasitol. 39:267-274 (1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92112918; PubMed=1370484;
 RA Ramamoorthy R., Donelson J.E., Paetz K.E., Maybodi M., Roberts S.C.,
 RA Wilson M.E.;
 RT "Three distinct RNAs for the surface protease gp63 are differentially
 RT expressed during development of Leishmania donovani chagasi
 RT promastigotes to an infectious form.";
 RL J. Biol. Chem. 267:1888-1895 (1992).
 CC -!- FUNCTION: Has an integral role during the infection of macrophages
 CC in the mammalian host.
 CC -!- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and
 CC P1' and basic residues at P2 and P3'. A model nonapeptide is
 CC cleaved at -Ala-Tyr-[Leu-Lys-Lys-
 CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
 CC -!- SIMILARITY: Belongs to peptidase family M8.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M80672; AAA29238.1; -.
 DR EMBL; M28527; AAA29235.1; -.
 DR PIR; A44951; A44951.
 DR HSSP; P08148; 1LML.
 DR MEROPS; M08.001; -.
 DR InterPro; IPR006025; Pept_M_Zn_BS.
 DR InterPro; IPR001577; Peptidase_M8.
 DR Pfam; PF01457; Peptidase_M8; 1.
 DR PRINTS; PRO0782; LSHMANOLYSIN.
 DR PROSITE; PS00142; ZINC_PROTEASE; 1.
 DR Hydrolase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;
 KW Zymogen; Signal; Cell adhesion; GPI-anchor; Lipoprotein.
 FT SIGNAL 1 39
 FT PROPEP 40 97
 FT CHAIN 98 574
 FT PROPEP 575 599
 FT METAL 261 261
 FT ACT_SITE 262 262
 FT METAL 265 265
 FT METAL 331 331
 FT DISULFID 122 139
 FT DISULFID 198 227
 FT DISULFID 311 383
 FT DISULFID 390 452
 FT DISULFID 403 422
 FT DISULFID 412 486
 FT DISULFID 463 507
 FT DISULFID 512 562
 FT DISULFID 532 555
 FT CARBOHYD 297 297
 FT CARBOHYD 394 394
 FT LIPID 574 574
 SQ SEQUENCE 599 AA; 63848 MW; 746730AE9E2A2E7C CRC64;
 Query Match 11.2%; Score 87; DB 1; Length 599;
 Best Local Similarity 89.5%; Pred. No. 4.4;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 108 AKNYDQLVTRVVTTHMAHA 126

DB 248 ASRYDQLVTRVVTTHMAHA 266

RESULT 13

GP63 LEIMA STANDARD; PET; 602 AA.
 ID GP63 LEIMA
 AC P08148; P15906;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Leishmanolysin precursor (EC 3.4.24.36) (Cell surface protease)
 DE (Major surface glycoprotein) (GP63 protein) (promastigote surface
 DE endopeptidase).
 GN GP63.
 OS Leishmania major.
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 OX NCBI_TaxID=5664;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 101-123.
 RX MEDLINE=88154764; PubMed=3346625;
 RA Button L.L., McMaster W.R.;
 RT "Molecular cloning of the major surface antigen of Leishmania.";
 RL J. Exp. Med. 167:724-729 (1988).
 RN [2]
 RP REVISIONS.
 RA Button L.L., McMaster W.R.;
 RL J. Exp. Med. 171:589-589 (1990).
 RN [3]
 RP GPI-ANCHOR.
 RX MEDLINE=91009116; PubMed=2145267;
 RA Schneider P., Ferguson M.A.J., McConville M.J., Mehlert A.,
 RA Homans S.W., Bordier C.;
 RT "Structure of the glycosyl-phosphatidylinositol membrane anchor of
 RT the Leishmania major promastigote surface protease.";
 RL J. Biol. Chem. 265:16955-16964 (1990).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).
 RX MEDLINE=95406217; PubMed=7675788;
 RA Schlegelhauf E., Etges R., Metcalf P.;
 RT "Crystallization and preliminary X-ray diffraction studies of
 RT Leishmanolysin, the major surface metalloproteinase from Leishmania
 RT major.";
 RL Proteins 22:58-66 (1995).
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY (1.86 ANGSTROMS).
 RX MEDLINE=98416698; PubMed=9739094;
 RA Schlegelhauf E., Etges R., Metcalf P.;
 RT "The crystal structure of the Leishmania major surface proteinase
 RT Leishmanolysin.";
 RL Structure 6:1035-1046 (1998).
 CC -!- FUNCTION: Has an integral role during the infection of macrophages
 CC in the mammalian host.
 CC -!- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and
 CC P1' and basic residues at P2 and P3'. A model nonapeptide is
 CC cleaved at -Ala-Tyr-[Leu-Lys-Lys-
 CC -!- COFACTOR: Binds 1 zinc ion per subunit.
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
 CC -!- PTM: THE PHOSPHATIDYLINOSITOL MOIETY OF THE GPI-ANCHOR CONTAINS A
 CC FULLY SATURATED, UNBRANCHED 1-O-ALKYL CHAIN (MAINLY C24:0) AND A
 CC MIXTURE OF FULLY SATURATED UNBRANCHED 2-O-ACYL CHAINS (C12:0,
 CC C14:0, C16:0, AND C18:0).
 CC -!- SIMILARITY: Belongs to peptidase family M8.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Y00647; CAA68673.1; -.
 DR PIR; PLO221; PLO221.
 DR PDB; 1LML; 17-SEP-97.
 DR MEROPS; M08.001; -.

FT	TURN	365	366
FT	STRAND	369	370
FT	HELIX	372	374
FT	TURN	380	383
FT	HELIX	386	390
FT	STRAND	394	395
FT	TURN	396	397
FT	STRAND	398	399
FT	TURN	402	404
FT	STRAND	413	414
FT	TURN	417	418
FT	STRAND	421	425
FT	STRAND	428	429
FT	HELIX	435	437
FT	TURN	443	444
FT	STRAND	445	446
FT	TURN	450	454
FT	STRAND	458	465
FT	TURN	466	467
FT	HELIX	470	472
FT	TURN	475	477
FT	HELIX	478	480
FT	TURN	485	486
FT	STRAND	487	494
FT	STRAND	496	496
FT	STRAND	506	516
FT	TURN	517	520
FT	STRAND	521	525
FT	TURN	527	528
FT	STRAND	533	534
FT	TURN	537	538
FT	STRAND	540	542
FT	HELIX	543	545
FT	TURN	546	546
FT	STRAND	550	550
FT	TURN	552	553
FT	STRAND	555	557
FT	HELIX	561	565
FT	TURN	566	567
FT	HELIX	569	572
FT	TURN	573	573
SEQ	SEQUENCE	602 AA; 63953 MW; 982EF3245D87C43E	CRC64;

Query Match 11.2%; Score 87; DB 1; Length 602;
 Best Local Similarity 89.5%; Pred. No. 4.5;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY	108 AKNYDQLVTRVVTHEMAHA	126
Db	251 ASRYDQLVTRVVTHEMAHA	269

RESULT 14
 GP63 LEIME
 ID GP63 LEIME STANDARD; PRT; 646 AA.
 AC P43150;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Leishmanolysin C1 precursor (EC 3.4.24.36) (Cell surface protease)
 DE DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface
 DE endopeptidase).
 GN GP63-C1
 OS Leishmania mexicana.
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 OX NCBI_TaxID=5665;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MNYC/BZ/62/M379;
 RC MEDLINE=93149206; PubMed=8426614;
 RA Medina-Acosta E., Kares R.E., Russell D.G.;
 RT "Structurally distinct genes for the surface protease of Leishmania
 RT mexicana are developmentally regulated.";

RL Mol. Biochem. Parasitol. 57:31-46(1993).
CC -!- FUNCTION: Has an integral role during the infection of macrophages
CC in the mammalian host.
CC -!- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and
CC P1' and basic residues at P2 and P3'. A model nonapeptide is
CC cleaved at -Ala-Tyr-|-Leu-Lys-Lys-.
CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC -!- DEVELOPMENTAL STAGE: Expressed in both the promastigote and the
CC amastigote forms.
CC -!- SIMILARITY: Belongs to peptidase family M8.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X64394; CAA45733.1; -.
DR PIR; S19916; S19916.
DR HSSP; P08148; 1LML.
DR MEROPS; M08.001; -.
DR GlycoSuiteDB; P43150; -.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR InterPro; IPR001577; Peptidase M8.
DR Pfam; PF01457; Peptidase M8; 1.
DR PRINTS; PR00782; LSHMANOLYSIN.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydrolase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;
KW Zymogen; Signal; Cell adhesion; Multigene family.
FT SIGNAL 1 39
FT PROPEP 40 102
FT CHAIN 103 646
FT METAL 266 266
FT ACT_SITE 267 267
FT METAL 270 270
FT METAL 336 336
FT DISULFID 127 144
FT DISULFID 193 232
FT DISULFID 316 388
FT DISULFID 395 458
FT DISULFID 408 427
FT DISULFID 417 492
FT DISULFID 469 513
FT DISULFID 518 568
FT DISULFID 538 561
FT CARBOHYD 86 86
FT CARBOHYD 297 297
FT CARBOHYD 399 399
FT CARBOHYD 409 409
FT CARBOHYD 433 433
FT CARBOHYD 445 445
FT CARBOHYD 466 466
FT CARBOHYD 501 501
FT CARBOHYD 501 501
SQ SEQUENCE 646 AA; 69054 MW; F448DDC78C10B0A CRC64;
Query Match 11.2%; Score 87; DB 1; Length 646;
Best Local Similarity 89.5%; Pred. No. 4.8;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 108 AKNYDQLVTRVVVTHEMAHA 126
DB 253 ASRDYDQLVTRVVVTHEMAHA 271
RESULT 15
ID OMPB_RICJA STANDARD; PRT; 1656 AA.
AC C06653;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Outer membrane protein B precursor (168 kDa surface-layer protein)
DE (Surface protein antigen) (Cell surface antigen 5) (Scas) (rOmpB)
DE [Contains: 120 kDa surface-exposed protein (Surface protein
DE antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide].
GN OMPB.
OS Rickettsia japonica.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsieae; Rickettsia.
OX NCBI_TaxID=35790;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YH;
RA Uchiyama T.;
RT "Sequencing of the gene encoding the protein rOmp B of Rickettsia
RT japonica";
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THE 120 kDa SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (By
CC similarity).
CC -!- FUNCTION: THE 32 kDa BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR
CC (By similarity).
CC -!- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-
CC layer with hexagonal symmetry.
CC -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB003681; BAA20138.1; -.
DR InterPro; IPR006345; Autotransport.
DR InterPro; IPR005546; Autotransporter.
DR Pfam; PF03797; Autotransporter; 1.
DR TIGRFAMs; TIGR01414; autotrans_bar1; 2.
KW Antigen; S-layer; Cell wall.
FT CHAIN 1 1338
FT CHAIN 1339 1656
FT DOMAIN 528 533
FT POLY-GLY.
SQ SEQUENCE 1656 AA; 168097 MW; 3132A69C9DD5999F CRC64;
Query Match 11.2%; Score 87; DB 1; Length 1656;
Best Local Similarity 29.6%; Pred. No. 14;
Matches 45; Conservative 14; Mismatches 55; Indels 38; Gaps 8;
QY 6 VAFAPAIIVSGSALAGVFPQWGGGNGHNGSSGDPSTLSIYQYGSANAALALQSDARK 65
DB 509 VLAAGAITLDGSATI-----TGDIGNGGG-----GAALQSITLANDATK 547
QY 66 SETTITQSG-----YGNQADVGGQADNSTIELTQNGERNNATIDQWAKNYDQLVTRVVTH 121
DB 548 ---TLTIGGANITISANGTINFGANGGTIKLTST---QNNIVVDCDLAIATDQ---TGVVDA 600
QY 122 EMAHANQTASDSWVRQVGF--GNNATANQY 151
DB 601 SSLTNAQTLTISGT-----IGIIGANNITLIGQF 628
Search completed: August 2, 2004, 14:49:31
Job time : 5.3 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:39:17 ; Search time 29.7 Seconds

(without alignments)
1604.150 Million cell updates/sec

Title: US-09-543-407-24

Perfect score: 775

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVGFNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	681	87.9	152	2	O33802 salmonella
2	591.5	76.3	150	2	Q7x243 citrobacter
3	553	71.4	149	2	Q7x240 citrobacter
4	508.5	65.6	152	16	Q8CW63 escherichia
5	431.5	55.7	150	2	Q7x237 enterobacte
6	385	49.7	76	2	Q54069 salmonella
7	122	15.7	29	2	Q9S3J5 escherichia
8	122	15.7	502	16	Q8E1H4 shewanella
9	109	14.1	139	16	Q8E1H3 shewanella
10	104	13.4	7716	16	Q7UW28 rhodospirillum
11	103	13.3	362	16	Q89D03 bradyrhizob
12	101	13.0	409	5	Q19414 caenorhabdi
13	100	12.9	151	2	Q7x238 enterobacte
14	100	12.9	362	16	Q8EV84 mycoplasma
15	99.5	12.8	171	16	Q89J13 bradyrhizob
16	99	12.8	1748	5	Q94821 tetrahymena

17	98	12.6	713	5	Q9N8N5	Q9n8n5 trypanosoma
18	96.5	12.5	151	16	Q7UCZ1	Q7uczi shigella fl
19	96.5	12.5	160	16	Q83RU7	Q83ru7 shigella fl
20	96	12.4	154	16	Q89J15	Q89j15 bradyrhizob
21	96	12.4	157	16	Q88HG0	Q88hg0 pseudomonas
22	95.5	12.3	151	2	Q7X244	Q7x244 citrobacter
23	95.5	12.3	160	16	Q8CM64	Q8cm64 escherichia
24	95.5	12.3	624	3	Q8N1V1	Q8n1v1 neurospora
25	95.5	12.3	1286	2	Q841Y5	Q841y5 campylobact
26	94.5	12.2	453	5	Q9N6M8	Q9n6m8 drosophila
27	94	12.1	179	2	Q33801	Q33801 salmonella
28	93.5	12.1	453	5	Q9NGF6	Q9ngf6 drosophila
29	93.5	12.1	453	5	Q9NGF7	Q9ngf7 drosophila
30	93.5	12.1	552	16	P96840	P96840 mycobacteri
31	93.5	12.1	552	16	Q7TW76	Q7tw76 mycobacteri
32	93.5	12.1	623	16	Q8V1Y0	Q8v1y0 mycobacteri
33	93	12.0	3501	16	Q8Y106	Q8y106 ralistonia s
34	93	12.0	3552	16	Q8XSD6	Q8xsd6 ralistonia s
35	92.5	11.9	348	13	Q93397	Q93397 cyprinus ca
36	92.5	11.9	440	16	Q92DG1	Q92dg1 listeria in
37	92.5	11.9	1222	5	Q9W4F0	Q9w4f0 drosophila
38	92.5	11.9	1222	5	Q8T8L9	Q8t8l9 drosophila
39	92.5	11.9	1351	5	Q8SX56	Q8sxs6 drosophila
40	92.5	11.9	1354	5	Q8MPN4	Q8mpn4 drosophila
41	91.5	11.8	1422	16	Q8EFU3	Q8efu3 shewanella
42	91.5	11.8	1615	2	Q9KKA8	Q9kka8 rickettsia
43	91	11.7	191	3	Q8TFA6	Q8tfa6 saccharomyc
44	90.5	11.7	152	2	Q7X241	Q7x241 citrobacter
45	90.5	11.7	1209	16	Q89CK5	Q89ck5 bradyrhizob

ALIGNMENTS

RESULT 1

O33802	PRELIMINARY;	PRT;	152 AA.
ID	O33802		
AC	O33802;		
DT	01-JAN-1998 (TrEMBLrel. 05, Created)		
DT	01-JAN-1998 (TrEMBLrel. 05, Last sequence update)		
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)		
DE	Agfa protein (fragment).		
GN	AGFA.		
OS	Salmonella typhimurium.		
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;		
OC	Enterobacteriaceae; Salmonella.		
OX	NCBI_TaxID=602;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=98033981; PubMed=9393832;		
RA	Sukupolvi S.S., Lorentz R.G., Gordon J.I., Bian Z., Pfeifer J.D.,		
RA	Normark S.J., Rhen M.;		
RT	"Expression of thin, aggregative fimbriae promotes interaction of		
RT	Salmonella typhimurium SR-11 with mouse small intestinal epithelial		
RT	cells.";		
RL	Infect. Immun. 65:5320-5325(1997).		
DR	EMBL; AJ000514; CAA04151.1; -		
FT	NON TER		
SQ	SEQUENCE 152 AA; 15401 MW; 9DAYDADC2364B006 CRC64;		

Query Match 87.9%; Score 681; DB 2; Length 152;
Best Local Similarity 90.1%; Pred. No. 2e-46;
Matches 136; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

Qy	1	MKLLKVAFAAIVVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQVGSANAALQ	60
Db	1	MKLLKVAFAAIVVSGSAVAGVVPQWGGGNNHNGGNSGPDSTLSIYQVGSANAALQ	60
Qy	61	SDARKSETTITQSGYNGADVCGAGNSITLTQNGFRNNTIDQNAKYDCLYTRVVT	120
Db	61	SDARKSETTITQSGYNGADVCGAGNSITLTQNGFRNNTIDQNAKYDCLYTRVVT	120
Qy	121	HEMAHANQTASDSSVMVRQVGFNNATANQY	151


```
RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RL Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158(2003).
DR EMBL; AJ515702; CAD56678.1; -.
SQ SEQUENCE 150 AA; 15112 MW; 5D8EB2D872DF15F3 CRC64;

Query Match 55.7%; Score 431.5; DB 2; Length 150;
Best Local Similarity 59.6%; Pred. No. 1e-26;
Matches 90; Conservative 27; Mismatches 33; Indels 1; Gaps 1;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQGGNGGNSGSPDSTLSIYQYGSANAALALQ 60
DQ :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1 MAFIKVAALAAIVVSGSAGAGINQ-GGWHGHGGGCGGPNSTLNIYQGGNSALALQ 59
QY 61 SPARKSETTITQSGYNGADVGQGDNDSTIETQNGFRNATIDQWNAKNDQLVTRVVT 120
DQ :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 60 TDARNVLNISQTGGNGADVGQGDSDSINTLTQNGFGNSATLDQWNSKDSVMNVSYQGG 119
QY 121 HEMAHANOTASDSSVMVROVGFNNATANQY 151
DQ :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 120 LAGALVDQTASNSTVNVTVQIFGNHATAHQY 150

RESULT 6
Q54069 PRELIMINARY; PRT; 76 AA.
AC Q54069;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SEF17 fimbria (Fragment).
GN AGFA.
OS Salmonella enteritidis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=592;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SE30;
RA COX J.M., Eglezos S., Woolcock J.B.;
RT "Virulence of Salmonella enteritidis in chickens correlates with
RT colony morphology and expression of SEF17 fimbriae.";
RL Submitted (APR-1996) to the EMBL/GenBank/DDJ databases.
DR EMBL; U53207; AAA98671.1; -.
FT NON_TER 1
FT NON_TER 76
SQ SEQUENCE 76 AA; 7704 MW; 2FD5411241A7BCB1 CRC64;

Query Match 49.7%; Score 385; DB 2; Length 76;
Best Local Similarity 98.7%; Pred. No. 2.1e-23;
Matches 75; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 30 GNHNGGGNSGSPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGADVGQGDNST 89
DQ :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1 GNHNGGGNSGSPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGADVGQGDNST 60
QY 90 IELTQNGFRNATIDQ 105
DQ :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 61 IELTQNGFRNATIDQ 75

RESULT 7
Q9S3J5 PRELIMINARY; PRT; 29 AA.
AC Q9S3J5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE Curlin subunit monomer (Fragment).
GN CSGA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
```

```
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC TRANSFOSON-insertion sequence ISI;
RX MEDLINE=99314153; PubMed=10386375;
RA La Ragione R.M., Collighan R.J., Woodward M.J.;
RT "Non-cultivation of Escherichia coli O78:X80 isolates associated with
RT ISI inserti on in csGB and reduced persistence in poultry infection.";
RL FEMS Microbiol. Lett. 175:247-253(1999).
DR EMBL; AJ131756; CAB45380.1; -.
FT NON_TER 29
FT NON_TER 29
SQ SEQUENCE 29 AA; 2789 MW; E290DFC07ABBB243 CRC64;

Query Match 15.7%; Score 122; DB 2; Length 29;
Best Local Similarity 89.7%; Pred. No. 0.0042;
Matches 26; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQGGG 29
DQ :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1 MKLLKVAALAAIVFSGSALAGVVPQYGGG 29

RESULT 8
Q8EIH4 PRELIMINARY; PRT; 502 AA.
AC Q8EIH4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Conserved hypothetical protein.
GN SO0865.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.
OX NCBI_TaxID=70863;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MR-1;
RX MEDLINE=2297666; PubMed=12368813;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
RA Meyer R.T., Tsapin A., Scott J., Beaman M., Brinkac L., Daugherty S.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
RA Vanathavan J., Weidman J., Imprim M., Lee K., Berry K., Lee C.,
RA Mueller J., Khouri H., Gill J., Utterback T.R., McDonald L.A.,
RA Feldblyum T.V., Smith H.O., Venter J.C., Neillson K.H., Fraser C.M.;
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis.";
RL Nat. Biotechnol. 20:1118-1123(2002).
DR EMBL; AE015332; AAN53941.1; -.
DR TIGR; SO0865; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 502 AA; 52441 MW; D08CA23D6C46B62D CRC64;

Query Match 15.7%; Score 122; DB 16; Length 502;
Best Local Similarity 27.0%; Pred. No. 0.12;
Matches 41; Conservative 22; Mismatches 57; Indels 32; Gaps 5;

QY 29 GGNHNG-----GGN-----SSGPDSTLSIYQYGSANA---ALALQS 61
DQ :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 231 GDNHTGTVYALAGSENDISMEQEGSNNTAYLSMTTGDNDTVDITQGDGNTVGDSTIADI 290
QY 62 DARKSETTITQSGYNGADVGQGDNDSTIETQNGFRNATIDQWNAKNDQLVTRVVT 121
DQ :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 291 QGDDNDITIKQKDSNGAEFQWGWDSNDVLDKQGDANFATGAYGTNDFDLSSKGDNN 350
QY 122 EMAHANQTASDSSVMVROVGFNGFN-----NATAN 149
DQ :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 351 ELV-APATGEDNSIEISQEGDANFAVVDATGN 381

RESULT 9
```

Best Local Similarity 29.1%; Pred. No. 81;
Matches 39; Conservative 18; Mismatches 57; Indels 20; Gaps 6;
QY 28 GCGNHN--GGNNS-----GPDSTLSIYQGSANAALQSDARKSETTITQSGY 75
DB 4048 GDNHGTGLGGVAVSQPTSSSVGEGDGLSV---SVGNGLANDIEMDGDLSITTEIN- 4103
QY 76 GNGADYVCGQ---ADNSTIETLQNGFERNATIOWNAKNYDQLVTRVVTHEMAHQAOTASD 132
DB 4104 GNSDYGNNITLPSGATIQLEADGSSYDPGTIYQNLNDGATATERTFTVSDGNGTDT 4163
QY 133 GSVWVRQVGFENNA 146
DB 4164 TSVAITIIG-GNDA 4176
RESULT 11
Q89D03 PRELIMINARY; PRT; 362 AA.
AC Q89D03;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical exported glutamine-rich protein.
GN BLU7642.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110.
RX MEDLINE=244998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA Sasamoto S., Watanabe A., Idesawa K., Iriguchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimpoto S., Tsuruoka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
Bradyrhizobium japonicum USDA110."
RL DNA Res. 9:189-197(2002).
DR EMBL; AP005962; BAC52907.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 362 AA; 39058 MW; D447EED04E8433EC CRC64;
Query Match 13.3%; Score 103; DB 16; Length 362;
Best Local Similarity 31.6%; Pred. No. 26;
Matches 50; Conservative 14; Mismatches 80; Indels 14; Gaps 6;
QY 2 KLLKVAFAAIVVSGSALAGVFPQWGGGHNHGGNSSGPDST----LSIYQGSANAAL 57
DB 3 KRLFLATTAATAVATSAQAQSSP---STNSNPSTTQRPDSTSTTPSSSTPSGSAQTNP 59
QY 58 ALQSDARKSETTITQSGYNGADVQGDADNSTIELTQNGFERNATI----DOWNAKNYDQ 113
DB 60 STNSAQTSPPSTGQSAAGQTTSGT-NTTAQTSNNSTNQASQPSQNTNAPS-DQ 117
QY 114 LVTRVVTHEMAH-ANQTSADSSVMVRQVGFENNAQ 150
DB 118 TQTPPTDNRQAQSANPPASGASQAQSPSGNSTNTAQ 155
RESULT 12
Q19414 PRELIMINARY; PRT; 409 AA.
AC Q19414;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE F13E9.4 protein.
GN F13E9.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.

Q8EIH3 PRELIMINARY; PRT; 139 AA.
AC Q8EIH3;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Minor curlin subunit CsgB, putative.
GN SO0866.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.
OX NCBI_TaxID=70863;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NR-1;
RX MEDLINE=2257686; PubMed=12368813;
RA Heidelberg J.F., Paulsen I.T., Nelson K.B., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
RA Vamathevan J., Weidman J., Umayam L.A., Lee K., Berry K., Lee C.,
RA Mueller J., Khouri H., Gill J., Utterback T.R., McDonald L.A., C.M.,
RA Feldblyum T.V., Smith H.O., Venter J.C., Nelson K.H., Fraser C.M.;
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
Shewanella oneidensis."
RL Nat. Biotechnol. 20:1118-1123(2002).
DR EMBL; AS015532; AAN53942.1; -
DR TIGR; SO0866; -
KW Complete proteome.
SQ SEQUENCE 139 AA; 14811 MW; 41ECLCPA76957920 CRC64;
Query Match 14.1%; Score 109; DB 16; Length 139;
Best Local Similarity 30.1%; Pred. No. 0.28;
Matches 34; Conservative 19; Mismatches 46; Indels 14; Gaps 3;
QY 39 GPDSTLSIYQGSANAALQSDARKSETTITQSGYNGADVQGDADNSTIELTQNGFR 98
DB 41 SGRDNLIDLQQQTQNGQVFGSGDNS-AYTQAGNDNISLVTOGTNNVOLLQVGAQ 99
QY 99 NNATIQWNAKNYDQLVTRVVTHEMAHQAOTASDSSVMVRQVGFENNAQ 151
DB 100 NKASITQIGNDNLVQL-----NQLGS-GNFSIQIADGAISITQY 139
RESULT 10
Q7UWZ8 PRELIMINARY; PRT; 7716 AA.
AC Q7UWZ8;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN RB1661.
OS Rhodospirillum rubrum.
OC Bacteria; Planctomycetes; Planctomycetacia; Planctomycetiales;
OC Planctomycetaceae; Pirellula.
OX NCBI_TaxID=117;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1;
RX MEDLINE=22735913; PubMed=12835416;
RA Glickner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,
RA Ludwig W., Gade D., Beck A., Borzym K., Heitmann K., Rabus R.,
RA Schlesner H., Amann R., Reinhardt R.;
RT "Complete genome sequence of the marine planctomycete Pirellula sp.
strain 1."
RT Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303(2003).
RL EMBL; BX294135; CAD72214.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 7716 AA; 797868 MW; D391A25BD96405C0 CRC64;
Query Match 13.4%; Score 104; DB 16; Length 7716;

```

OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA McMurray A.A.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
   none;
RA "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL: Z69383; CAA93412.1; -.
DR PIR: T20847; T20847.
DR WormPep: F13E9.4; CE05606.
DR InterPro: IPR003677; Onchoerca_Ag.
DR Pfan; PF02520; DUF148; 1.
SQ SEQUENCE 409 AA; 43231 MW; B07DFOE4175C5739 CRC64;

Query Match      13.0%; Score 101; DB 5; Length 409;
Best Local Similarity 23.3%; Pred. No. 4.3;
Matches 37; Conservative 26; Mismatches 65; Indels 30; Gaps 5;

QY 15 SGSALAGVVPWGGGNGGNSGSDSTLSIYQVGSANAALALQSDARKSETTITQSG 74
DQ 134 SGQAGSGMNFPGGGGQYGNQNGFGGQSGGSGGNSLSANSNGNNQG--SSSG 191

QY 75 YGN--GADVGGGANSTIE---LTQNGFRNN-----ATIDQNNAX 109
DQ 192 YQNNQGRHQGGGSHSSNSVMSNGYSSNGYNNNGPTPSPFLNNVSSSAQDYTNIV 251

QY 110 NYDQLVTRVTHMAHANQTSDDSVVMVQVGFGNATA 148
DQ 252 NKSLTTNQINEQ---ASNWASNSVQAYIQYETNRSA 287

RESULT 13
Q7X238 PRELIMINARY; PRT; 151 AA.
AC Q7X238
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DE Nucleation component of curlin monomers.
GN CSGB.
OS Enterobacter sakazakii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Enterobacter.
OX NCBI_TaxID=28141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fec39;
RA Zogaj X., Bokranz W., Nimtz M., Romling U.;
RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RT Enterobacteriaceae isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158(2003).
DR EMBL: AJ515702; CAD56677.1; -.
SQ SEQUENCE 151 AA; 15985 MW; F0B82BD2A27882B7 CRC64;

Query Match      12.9%; Score 100; DB 2; Length 151;
Best Local Similarity 31.8%; Pred. No. 1.6;
Matches 35; Conservative 12; Mismatches 45; Indels 18; Gaps 4;

QY 10 AAIVVGSALAGVVPQWGGGNGGNSGSDSTLSIYQVGSANAALALQSDARKSETT 69
DQ 58 AQIRQSGKLLSVVQ-----DGAGNRAEVD-----QSGTYNIWIDQS--GNGNDAG 103

QY 70 ITQSGYNGADVGGGANSTIELTQNGFRNNATTDQNNAKNYDQLVTRVV 119
DQ 104 ITQDGYGNSAKIIQKSGNSRANITQYGTAVVVQ-----KQSQMAIRVI 149

RESULT 14

```

```

Q8EV84 PRELIMINARY; PRT; 362 AA.
AC Q8EV84
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE P35 lipoprotein homolog.
GN MYPE6840.
OS Mycoplasma penetrans.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=26227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HF-2;
RX MEDLINE=22354719; PubMed=12465555;
RA Sasaki Y., Ishikawa J., Yamashita A., Oshima K., Kenri T., Furuya K.,
RA Yoshino C., Horino A., Shiba T., Sasaki T., Hattori M.;
RT "The complete genomic sequence of Mycoplasma penetrans, an
RT intracellular bacterial pathogen in humans.";
RL Nucleic Acids Res. 30:5293-5300(2002).
DR EMBL: AF004172; BAC44476.1; -.
DR InterPro: IPR000437; Prok Lipoprot S.
DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
KW Lipoprotein; Complete proteome.
SQ SEQUENCE 362 AA; 38547 MW; 8DA27F70D19D354F CRC64;

Query Match      12.9%; Score 100; DB 16; Length 362;
Best Local Similarity 23.9%; Pred. No. 4.5;
Matches 42; Conservative 33; Mismatches 63; Indels 38; Gaps 8;

QY 1 MLLKVAAPAAIVVSGS-ALAGVVP-----QWGGGNGHNGGNSG-----PDS 43
DQ 1 MKIKKILKALALTGAFIVATVPVIVSSCSTSDNNGNGNNNNNGQDGGGQOOTEI 60

QY 44 TLSIYQVGSANAALALQSDARKSETTITQSG-----YNGADYGGQGANSTIEL--- 92
DQ 61 TPTIKKEVSLSGALSKIYDANKSTSLIAEDIKANPTNYDFNGEALKDLIKDATVSVNG 120

QY 93 -TQNGFRNNATTDQNNAKNYDQLVTRVTHMAHANQTSDDSV--MVRQVGFGN 145
DQ 121 FTESTFKGD-TVETWSAKYGVKKGT-----YQAASKQLDIKSINDLETQLGDSNN 169

RESULT 15
Q89J13 PRELIMINARY; PRT; 171 AA.
AC Q89J13
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DE CsgA protein.
GN CSGA OR BLJ5300.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RX MEDLINE=22484998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA Sasamoto S., Watanabe A., Idesawa K., Iriguchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimpo S., Tsurucka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110.";
RL DNA Res. 9:189-197(2002).
DR EMBL: AP005954; BAC50565.1; -.
KW Complete proteome.
SQ SEQUENCE 171 AA; 17448 MW; 995DB08C01498381 CRC64;

Query Match      12.8%; Score 99.5; DB 16; Length 171;
Best Local Similarity 25.8%; Pred. No. 2;

```

Matches	40;	Conservative	21;	Mismatches	63;	Indels	31;	Gaps	4;
QY	1	MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ	60						
Db	40	MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ	60						
QY	61	SDARKSETTITQSGYNGADVGGAD---	116						
Db	73	NGLTNDSSSTTQIGILNGASTMQGTSPSLNNVSTVNOAGVQNSATTGQVAFGNGSAIT	138						
QY	117	RVVTHEMAHANOTASDSSVMVQVGF--NNATANQ	150						
Db	139	QNSFGPPALQNNAS-----VGQLSFGINTSTVSQ	168						

Search completed: August 2, 2004, 14:54:40
Job time : 29.7 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:35:42 ; Search time 44.9 Seconds

(without alignments)
950.215 Million cell updates/sec

Title: US-09-543-407-26

Perfect score: 782

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVGFNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : A.Geneseq 29Jan04:*

- 1: geneseqp1980s:*
- 2: geneseqp1990s:*
- 3: geneseqp2000s:*
- 4: geneseqp2001s:*
- 5: geneseqp2002s:*
- 6: geneseqp2003as:*
- 7: geneseqp2003bs:*
- 8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	782	100.0	151	3 AAB36353	Aab36353 Agfa::PT3
2	709	90.7	151	3 AAB36350	Aab36350 Agfa::PT3
3	692	88.5	151	2 AAR74625	Aar74625 Agfa sequ
4	692	88.5	151	3 AAB36341	Aab36341 Salmonell
5	687	87.9	151	2 AAW23570	Aaw23570 Salmonell
6	675	86.3	151	3 AAB36349	Aab36349 Agfa::PT3
7	617	78.9	151	3 AAB36354	Aab36354 Agfa::PT3
8	614	78.5	151	3 AAB36346	Aab36346 Agfa::PT3
9	612	78.3	151	3 AAB36347	Aab36347 Agfa::PT3
10	609	77.9	151	3 AAB36352	Aab36352 Agfa::PT3
11	601	76.9	151	3 AAB36351	Aab36351 Agfa::PT3
12	600	76.7	151	3 AAB36355	Aab36355 Agfa::PT3
13	577	73.8	151	3 AAB36348	Aab36348 Agfa::PT3
14	523	66.9	151	3 AAB36343	Aab36343 Escherich
15	518	66.2	151	7 ABR82651	AbR82651 E. coli C
16	507	64.8	120	2 AAR62761	Aar62761 Agfa sequ
17	507	64.8	120	2 AAW23569	Aaw23569 Salmonell
18	445	56.9	142	2 AAR52864	Aar52864 Fibronect
19	373	47.7	122	2 AAR52863	Aar52863 FNB curli
20	188	24.0	45	3 AAB36316	Aab36316 Salmonell
21	132	16.9	22	3 AAB36318	Aab36318 Salmonell
22	123	15.7	23	3 AAB36321	Aab36321 Salmonell
23	123	15.7	23	3 AAB36326	Aab36326 Salmonell
24	123	15.7	23	3 AAB36338	Aab36338 Salmonell
25	113	14.5	24	7 ABR82844	AbR82844 E. coli c

26	112	14.3	151	3 AAB36344	Aab36344 Escherich
27	111	14.2	22	3 AAB36322	Aab36322 Salmonell
28	111	14.2	22	3 AAB36327	Aab36327 Salmonell
29	111	14.2	22	3 AAB36337	Aab36337 Salmonell
30	107	13.7	151	3 AAB36342	Aab36342 Salmonell
31	105	13.4	597	4 AAB08231	Aau08231 Polypepti
32	102	13.0	26	7 ABR82649	AbR82649 E. coli V
33	100	12.8	943	2 AAW64378	Aaw64378 Mycobacte
34	100	12.8	943	2 AAW61745	Aaw61745 M. tuberc
35	100	12.8	943	2 AAY39032	Aay39032 M. tuberc
36	100	12.8	943	2 AAY39175	Aay39175 M. tuberc
37	100	12.8	3300	6 ABU36445	Abu36445 Protein e
38	98.5	12.6	738	2 AAW56163	Aaw56163 New DNA s
39	97.5	12.5	850	4 ABB65764	Abb65764 Drosophil
40	97.5	12.5	1028	4 ABB62708	Abb62708 Drosophil
41	96.5	12.3	447	3 AAG29728	Aag29728 Arabidops
42	96.5	12.3	468	3 AAG29727	Aag29727 Arabidops
43	96	12.3	19	3 AAB36323	Aab36323 Salmonell
44	96	12.3	19	3 AAB36336	Aab36336 Salmonell
45	96	12.3	19	3 AAB36328	Aab36328 Salmonell

ALIGNMENTS

RESULT 1
AAB36353
ID AAB36353 standard; protein; 151 AA.
XX
AC AAB36353;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa::PT3#8 amino acid sequence SEQ ID NO:26.
XX
KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO2000060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UUVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
XX
N-PSDB; AAC64629.
XX
PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 138; 139pp; English.
XX
CC The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEFI7/7AF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native

CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fibrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;

Query Match 100.0%; Score 782; DB 3; Length 151;
 Best Local Similarity 100.0%; Pred. No. 2.3e-67;
 Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MLLKVAAPAAIVVSGSALAGVVPQWGGGNGGNGSGGPDSTLSIYQGSANAALYDQ 60
 DB 1 MLLKVAAPAAIVVSGSALAGVVPQWGGGNGGNGSGGPDSTLSIYQGSANAALYDQ 60
 QY 61 LVTRVVTHEMAHAGYNGADVGQGDNSTIETQNGFRNNATIDOWNAKNSDITVGQYGG 120
 DB 61 LVTRVVTHEMAHAGYNGADVGQGDNSTIETQNGFRNNATIDOWNAKNSDITVGQYGG 120
 QY 121 NNAAALVNQTASDSSVMVRQVGFNNATANQY 151
 DB 121 NNAAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 2
 AAB36350
 ID AAB36350 standard; protein; 151 AA.
 XX
 AC AAB36350;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Agfa::P3#5 amino acid sequence SEQ ID NO:20.
 XX
 KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
 XX vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 FN WO2000060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 XX 05-APR-2000; 2000WO-CA000356.
 PF
 XX 05-APR-1999; 99US-0127888P.
 PR
 XX (UYVI-) UNIV VICTORIA.
 PA
 XX White AP, Doran JL, Collison SK, Kay WW;
 PI
 XX WPI; 2000-672631/65.
 DR N-PSDB; AAC64626.
 DR
 XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 XX Disclosure; Page 137; 139pp; English.
 PS

XX The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAf) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fibrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;

Query Match 90.7%; Score 709; DB 3; Length 151;
 Best Local Similarity 89.9%; Pred. No. 2.5e-60;
 Matches 143; Conservative 0; Mismatches 0; Indels 16; Gaps 2;
 QY 1 MLLKVAAPAAIVVSGSALAGVVPQWGGGNGGNGSGGPDSTLSIYQGSANAAL--- 57
 DB 1 MLLKVAAPAAIVVSGSALAGVVPQWGGGNGGNGSGGPDSTLSIYQGSANAALAQ 60
 QY 58 -----YDOLVTRVVTHEMAHAGYNGADVGQGDNSTIETQNGFRNNATIDOWNAKNSD 112
 DB 61 SDARKYDQLVTRVVTHEMAHA-----GQGDNSTIETQNGFRNNATIDOWNAKNSD 112
 QY 113 ITVGQYGGNNAALVNQTASDSSVMVRQVGFNNATANQY 151
 DB 113 ITVGQYGGNNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 3
 AAR74625
 ID AAR74625 standard; protein; 151 AA.
 XX
 AC AAR74625;
 XX
 DT 25-MAR-2003 (revised)
 DT 26-JUN-1995 (first entry)
 XX
 DE Agfa sequence.
 XX
 KW Salmonella; Agfa; vaccine.
 XX
 OS Salmonella.
 XX
 FN WO9425598-A2.
 PD 10-NOV-1994.
 XX
 XX 26-APR-1994; 94WO-IB000207.
 PF
 XX 26-APR-1993; 93US-00054452.
 PR
 XX (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
 PA (KING/) KING J.
 XX

PI Kay WW, Collinson SK, Clouthier SC, Doran JL;
 XX WPI; 1994-358275/44.
 DR N-PSDB; AAC87467.
 XX
 PT Eliciting an immune response to Salmonella - using attenuated Salmonella
 strains, vector constructs, or compenss. contg. fimbrial type proteins.
 XX
 PS Disclosure; Fig 7B; 95pp; English.
 XX
 CC The Salmonella Agfa protein and DNA are used in vaccine and genetic
 CC immunization compositions, respectively, to elicit an immune response to
 CC Salmonella in animals (e.g. food producing animals) and humans. (Updated
 CC on 25-MAR-2003 to correct FN field.)
 XX
 SQ Sequence 151 AA;
 Query Match 88.5%; Score 692; DB 2; Length 151;
 Best Local Similarity 90.7%; Pred. No. 1.1e-58;
 Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQYGSANAALYDQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQYGSANAALQ 60
 QY 61 LVTRVVTHEMAHAGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
 DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
 QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
 DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
 RESULT 4
 AAB36341
 ID AAB36341 standard; protein; 151 AA.
 AC AAB36341;
 DT 26-FEB-2001 (first entry)
 DE Salmonella enteritidis Agfa amino acid sequence SEQ ID NO:5.
 KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 PN WO200060102-A2.
 PD 12-OCT-2000.
 PF 05-APR-2000; 2000WO-CA000356.
 PR 05-APR-1999; 99US-0127888P.
 XX (UYVI-) UNIV VICTORIA.
 XX White AP, Doran JL, Collinson SK, Kay WW;
 DR WPI: 2000-672631/65.
 DR N-PSDB; AAC64617.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 135; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:

CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;
 Query Match 88.5%; Score 692; DB 3; Length 151;
 Best Local Similarity 90.7%; Pred. No. 1.1e-58;
 Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQYGSANAALYDQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQYGSANAALQ 60
 QY 61 LVTRVVTHEMAHAGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
 DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
 QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
 DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
 RESULT 5
 AAW23570
 ID AAW23570 standard; protein; 151 AA.
 AC AAW23570;
 DT 25-MAR-2003 (revised)
 DT 29-SEP-1997 (first entry)
 DE Salmonella enteritidis 2765S-3b agfa.
 XX
 KW Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody.
 XX Salmonella enteritidis.
 XX Key Location/Qualifiers
 FH Misc-difference 123
 FT /note= "Encoded by GCC"
 FT
 XX US5635617-A.
 PN
 XX 03-JUN-1997.
 PD
 XX 26-APR-1994; 94US-00233788.
 PF
 XX 26-APR-1993; 93US-00054452.
 PR
 XX (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
 PA
 XX Collinson SK, Kay WW, Doran JL;
 PI

```

XX WPI; 1997-309886/28.
XX N-PSDB; AAT74142.
XX
XX Isolated Salmonella gene agfA - used for diagnosis of Salmonella or
XX enteropathogenic bacteria of the Enterobacteriaceae family.
XX
XX Example 2; Fig 7; 85pp; English.
XX
XX The present sequence represents agfA encoded by the full agfA gene
XX derived from Salmonella enteritidis 27655-3b. The nucleic acid can be
XX used to provide diagnostic assays for Salmonella and/or enteropathogenic
XX bacteria of the family Enterobacteriaceae. It can also be used to provide
XX proteins and antibodies which can be used for assays. The nucleic acid
XX sequence can be used to provide probes or primers which can specifically
XX hybridize to nucleic acid molecules from greater than 99% of Salmonella
XX strains that are pathogenic to warm-blooded animals relative to nucleic
XX acid molecules from virtually all other microbial organisms. (Updated on
XX 25-MAR-2003 to correct PF field.)
XX
XX Sequence 151 AA;
SQ
Query Match      87.9%; Score 687; DB 2; Length 151;
Best Local Similarity 90.1%; Pred. No. 3.3e-58;
Matches 136; Conservative 2; Mismatches 13; Indels 0; Gaps 0;
QY 1 M K L L K V A A F A A I V V G S A L A G V V P Q W G G G N H N G G N S G P D S T L S I Y Q Y G S A N A A L Y D Q 60
D b 1 M K L L K V A A F A A I V V G S A L A G V V P Q W G G G N H N G G N S G P D S T L S I Y Q Y G S A N A A L A L Q 60
QY 61 L V T R V V T H E M A G Y G N G A D V G G A D N S T I E L T Q N G F R N N A T I D Q W N A K N S D I T V G Y G G 120
D b 61 S D A R K S E T T I T Q G Y G N G A D V G G A D N S T I E L T Q N G F R N N A T I D Q W N A K N S D I T V G Y G G 120
QY 121 N N A A L V N Q T A S D S S V M V R Q V G F G N N A T A N Q Y 151
D b 121 N N P A L V N Q T A S D S S V M V R Q V G F G N N A T A N Q Y 151
RESULT 6
AAB36349
ID AAB36349 standard; protein; 151 AA.
AC AAB36349;
XX
XX 26-FEB-2001 (first entry)
DE
DE AgfA::PT3#4 amino acid sequence SEQ ID NO:18.
XX
XX Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen.
XX
XX Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
XX WO200060102-A2.
FN
XX
XX 12-OCT-2000.
PD
XX
XX 05-APR-2000; 2000WO-CA000356.
PF
XX
XX 05-APR-1999; 99US-0127888P.
PR
XX
XX (UYVI-) UNIV VICTORIA.
FA
XX
XX White AP, Doran JL, Collison SK, Kay WW;
PI
XX
XX WPI; 2000-672631/65.
DR
XX
XX N-PSDB; AAC64625.
DR
XX
XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence
XX which encodes foreign epitope or antigen, expresses recombinant AgfA

```

```

PT protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 136; 139pp; English.
XX
XX The present invention describes a recombinant agfA gene (I) where a
XX segment of the gene has been replaced by a segment of a foreign DNA
XX sequence which encodes a foreign epitope or antigen. Also described are:
XX (1) use of thin aggregative fimbriae (SEF7/7AF) nucleation depended
XX assembly system of strains of Salmonella, Escherichia coli and
XX Enterobacteriaceae for the production of fimbriae comprising recombinant
XX AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
XX directing recombination of a recombinant gene into the chromosome of the
XX homologous species; (3) directing recombination of a recombinant gene
XX back into the chromosome of the homologous species, replacing the native
XX copy of that gene; and (4) eliciting an immune response in an animal, AgfA
XX comprising separating an amino acid polymer comprising a recombinant amino
XX protein containing a replacement segment or segments of foreign amino
XX acid sequence or sequences grown on a Salmonella, E. coli or
XX Enterobacteriaceae host cell, from the host cell and introducing the
XX polymer into the animal in conjunction with a carrier or diluent. (I) is
XX useful for the expression of recombinant AgfA protein which is useful for
XX eliciting an immune response in an animal. In a fimbrial presentation
XX system the heterologous antigens are presented in high numbers (up to
XX 500,000 copies/cell), the hybrid fimbrial protein possesses both the
XX immunogenicity and adhesion properties relevant for an efficient live
XX vaccine, the carrier fimbrial subunit proteins are usually strong
XX immunogens, which may be important for directing an immune response
XX against the inserted epitope, and hybrid fimbriae are easy and
XX inexpensive to purify in large amount. The present sequence is given in
XX the exemplification of the present invention
XX
XX Sequence 151 AA;
Query Match      86.3%; Score 675; DB 3; Length 151;
Best Local Similarity 81.9%; Pred. No. 4.7e-57;
Matches 136; Conservative 0; Mismatches 0; Indels 30; Gaps 2;
QY 1 M K L L K V A A F A A I V V G S A L A G V V P Q W G G G N H N G G N S G P D S T L S I Y Q Y G S A N A A L Y D Q 60
D b 1 M K L L K V A A F A A I V V G S A L A G V V P Q W G G G N H N G G N S G P D S T L S I Y Q Y G S A N A A L Y D Q 60
QY 61 L V T R V V T H E M A H A -----GYGNGADYGGADNSITELTQNGFRNNATIDQ 105
D b 46 L V T R V V T H E M A L Q S D A R K S E T T I T Q S G Y G N G A D V G G A D N S T I E L T Q N G F R N N A T I D Q 105
QY 106 W N A K N S D I T V G Y G G N N A A L V N Q T A S D S S V M V R Q V G F G N N A T A N Q Y 151
D b 106 W N A K N S D I T V G Y G G N N A A L V N Q T A S D S S V M V R Q V G F G N N A T A N Q Y 151
RESULT 7
AAB36354
ID AAB36354 standard; protein; 151 AA.
XX
XX AAB36354;
XX
XX 26-FEB-2001 (first entry)
DT
XX
XX AgfA::PT3#9 amino acid sequence SEQ ID NO:28.
DE
XX
XX Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen.
XX
XX Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
XX WO200060102-A2.
FN
XX
XX 12-OCT-2000.
PD
XX
XX 05-APR-2000; 2000WO-CA000356.
PF
XX
XX
XX

```

```
PR 05-APR-1999; 99US-0127888P.
XX (UYVI-) UNIV VICTORIA.
XX White AP, Doran JL, Collison SK, Kay WW;
XX WPI: 2000-672631/65.
XX N-PSDB; AAC64630.
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 138; 139pp; English.
XX
XX The present invention describes a recombinant agfa gene (I) where a
XX segment of the gene has been replaced by a segment of a foreign DNA
XX sequence which encodes a foreign epitope or antigen. Also described are:
XX (1) use of thin aggregative fimbriae (SEF17/TAf) nucleation depended
XX assembly system of strains of Salmonella, Escherichia coli and
XX Enterobacteriaceae for the production of fimbriae comprising recombinant
XX Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
XX directing recombination of a recombinant gene into the chromosome of the
XX homologous species; (3) directing recombination of a recombinant gene
XX back into the chromosome of the homologous species, replacing the native
XX copy of that gene; and (4) eliciting an immune response in an animal,
XX comprising separating an amino acid polymer comprising a recombinant Agfa
XX protein containing a replacement segment or segments of foreign amino
XX acid sequence or sequences grown on a Salmonella, E. coli or
XX Enterobacteriaceae host cell, from the host cell and introducing the
XX polymer into the animal in conjunction with a carrier or diluent. (I) is
XX useful for the expression of recombinant Agfa protein which is useful for
XX eliciting an immune response in an animal. In a fimbrial presentation
XX system the heterologous antigens are presented in high numbers (up to
XX 500,000 copies/cell), the hybrid fimbrial protein possesses both the
XX immunogenicity and adhesion properties relevant for an efficient live
XX vaccine, the carrier fimbrial subunit proteins are usually strong
XX immunogens, which may be important for directing an immune response
XX against the inserted epitope, and hybrid fimbriae are easy and
XX inexpensive to purify in large amount. The present sequence is given in
XX the exemplification of the present invention
XX
XX Sequence 151 AA;
XX
XX Query Match 78.9%; Score 61.7; DB 3; Length 151;
XX Best Local Similarity 73.6%; Pred. No. 1.8e-51;
XX Matches 128; Conservative 0; Mismatches 0; Indels 46; Gaps 2;
XX
XX QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQGSANAALQ 57
XX Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQGSANAALQ 60
XX
XX QY 58 -----YDQLVTRVVTHEMAHAGYNGADVGQGDNSTIETQNGF 97
XX Db 61 SDARKSETTITQSGYNGADVDQLVTRVVTHEMAHA-----F 97
XX
XX QY 98 RNNATIDQWNAKNSDITVQGYGNNALVNQTASDSSVMVRQVGFNNATANQY 151
XX Db 98 RNNATIDQWNAKNSDITVQGYGNNALVNQTASDSSVMVRQVGFNNATANQY 151
XX
XX RESULT 8
XX AAB36346
XX ID AAB36346 standard; protein; 151 AA.
XX AC AAB36346;
XX XX
XX DT 26-FEB-2001 (first entry)
XX XX
XX DE Agfa::PT#1 amino acid sequence SEQ ID NO:12.
XX KW Salmonella; agfa; chromosomal gene replacement; fimbrial; epitope;
XX vaccine; immune response; immunogen.
```

```
XX Salmonella enteritidis.
XX Escherichia coli.
XX Synthetic.
XX W0200060102-A2.
XX 12-OCT-2000.
XX 05-APR-2000; 2000WC-CA000356.
XX 05-APR-1999; 99US-0127888P.
XX (UYVI-) UNIV VICTORIA.
XX White AP, Doran JL, Collison SK, Kay WW;
XX WPI: 2000-672631/65.
XX N-PSDB; AAC64622.
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 135; 139pp; English.
XX
XX The present invention describes a recombinant agfa gene (I) where a
XX segment of the gene has been replaced by a segment of a foreign DNA
XX sequence which encodes a foreign epitope or antigen. Also described are:
XX (1) use of thin aggregative fimbriae (SEF17/TAf) nucleation depended
XX assembly system of strains of Salmonella, Escherichia coli and
XX Enterobacteriaceae for the production of fimbriae comprising recombinant
XX Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
XX directing recombination of a recombinant gene into the chromosome of the
XX homologous species; (3) directing recombination of a recombinant gene
XX back into the chromosome of the homologous species, replacing the native
XX copy of that gene; and (4) eliciting an immune response in an animal,
XX comprising separating an amino acid polymer comprising a recombinant Agfa
XX protein containing a replacement segment or segments of foreign amino
XX acid sequence or sequences grown on a Salmonella, E. coli or
XX Enterobacteriaceae host cell, from the host cell and introducing the
XX polymer into the animal in conjunction with a carrier or diluent. (I) is
XX useful for the expression of recombinant Agfa protein which is useful for
XX eliciting an immune response in an animal. In a fimbrial presentation
XX system the heterologous antigens are presented in high numbers (up to
XX 500,000 copies/cell), the hybrid fimbrial protein possesses both the
XX immunogenicity and adhesion properties relevant for an efficient live
XX vaccine, the carrier fimbrial subunit proteins are usually strong
XX immunogens, which may be important for directing an immune response
XX against the inserted epitope, and hybrid fimbriae are easy and
XX inexpensive to purify in large amount. The present sequence is given in
XX the exemplification of the present invention
XX
XX Sequence 151 AA;
XX
XX Query Match 78.5%; Score 61.4; DB 3; Length 151;
XX Best Local Similarity 80.8%; Pred. No. 3.5e-51;
XX Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;
XX
XX QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQGSANAALQ 60
XX Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQGSANAALQ 60
XX
XX QY 61 LVTRVVTHEMAHAGYNGADVGQGDNSTIETQNGFRNNATIDQWNAKNSDITVQYGG 120
XX Db 61 SDARKSETTITQSGYNGADVGQGDNSTIETQNGFRNNATIDQWNAKNSDITVQYGG 120
XX
XX QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
XX Db 121 NNAALVNQDQLVTRVVTHEMAHANNATANQY 151
XX
XX RESULT 9
```

AAB36347
 ID AAB36347 standard; protein; 151 AA.
 AC AAB36347;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Agfa::PT3#2 amino acid sequence SEQ ID NO:14.
 XX
 KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 PN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay WW;
 XX
 DR WPI; 2000-672631/65.
 DR N-PSDB; AAC64623.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 136; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbriin subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbriin protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;
 Query Match 78.3%; Score 612; DB 3; Length 151;
 Best Local Similarity 81.5%; Pred. No. 5.5e-51;
 Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGGPDSTLSIYQGSANAALYDQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGGPDSTLSIYQGSANAALALQ 60

61 LVTRVVTHEMAHAGYNGADVGQADNSTIELTQNGFRNATIDQWNAKNSDITVQYGG 120
 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNATIDQWNAKNSDITVQYDQ 120
 121 NNAALVNQTASDSSVMVRQVGFQNNATANQY 151
 121 LVTRVVTHEMAHASVMVRQVGFQNNATANQY 151

RESULT 10
 AAB36352
 ID AAB36352 standard; protein; 151 AA.
 AC AAB36352;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Agfa::PT3#7 amino acid sequence SEQ ID NO:24.
 XX
 KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 PN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay WW;
 XX
 DR WPI; 2000-672631/65.
 DR N-PSDB; AAC64628.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 138; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbriin subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbriin protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

CC the exemplification of the present invention
XX Sequence 151 AA;
SQ

Query Match 77.9%; Score 609; DB 3; Length 151;
Best Local Similarity 82.1%; Pred. No. 1.1e-50;
Matches 124; Conservative 4; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALYDQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALQ 60
QY 61 LVTRVVTHEMAHAGYNGADYQGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGOYGG 120
DB 61 SDARKSETTITQSGYNGADYQGGADNSTIELTQNGFRNNATIDQWNAKNDYQLVTRVVT 120
QY 121 NNAALVNOTASDSSVMVQVGFNNATANQY 151
DB 121 HEMAHANGTASDSSVMVQVGFNNATANQY 151

RESULT 11
AAB36351
ID AAB36351 standard; protein; 151 AA.
XX
AC AAB36351;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#6 amino acid sequence SEQ ID NO:22.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
DN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UUVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI: 2000-672631/65.
XX
PS Disclosure; Page 137; 139pp; English.

The present invention describes a recombinant agfA gene (I) where a segment of the gene has been replaced by a segment of a foreign DNA sequence which encodes a foreign epitope or antigen. Also described are:
(1) use of thin aggregative fimbriae (SEF17/7AF) nucleation depended assembly system of strains of Salmonella, Escherichia coli and Enterobacteriaceae for the production of fimbriae comprising recombinant AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2) directing recombination of a recombinant gene into the chromosome of the homologous species; (3) directing recombination of a recombinant gene back into the chromosome of the homologous species, replacing the native copy of that gene; and (4) eliciting an immune response in an animal, comprising separating an amino acid polymer comprising a recombinant AgfA protein containing a replacement segment or segments of foreign amino acid sequence or sequences grown on a Salmonella, E. coli or

Enterobacteriaceae host cell, from the host cell and introducing the polymer into the animal in conjunction with a carrier or diluent. (I) is useful for the expression of recombinant AgfA protein which is useful for eliciting an immune response in an animal. In a fimbrial presentation system the heterologous antigens are presented in high numbers (up to 500,000 copies/cell), the hybrid fimbrial protein possesses both the immunogenicity and adhesion properties relevant for an efficient live vaccine, the carrier fimbrial subunit proteins are usually strong immunogens, which may be important for directing an immune response against the inserted epitope, and hybrid fimbriae are easy and inexpensive to purify in large amount. The present sequence is given in the exemplification of the present invention

Query Match 76.9%; Score 601; DB 3; Length 151;
Best Local Similarity 81.5%; Pred. No. 6.3e-50;
Matches 123; Conservative 4; Mismatches 24; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALYDQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALQ 60
QY 61 LVTRVVTHEMAHAGYNGADYQGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGOYGG 120
DB 61 SDARKSETTITQSGYNGADYQGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGOYGG 120
QY 121 NNAALVNOTASDSSVMVQVGFNNATANQY 151
DB 121 NNAALVNOTASDSSVMVQVGFNNATANQY 151

RESULT 12
AAB36355
ID AAB36355 standard; protein; 151 AA.
XX
AC AAB36355;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#10 amino acid sequence SEQ ID NO:30.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
DN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UUVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI: 2000-672631/65.
XX
PS Disclosure; Page 139; 139pp; English.

The present invention describes a recombinant agfA gene (I) where a segment of the gene has been replaced by a segment of a foreign DNA sequence which encodes a foreign epitope or antigen. Also described are:

CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention

XX Sequence 151 AA;

Query Match 76.7%; Score 600; DB 3; Length 151;
Best Local Similarity 81.5%; Pred. No. 7.8e-50;
Matches 123; Conservative 4; Mismatches 24; Indels 0; Gaps 0;
QY 1 MKLLKVAAPAAIVVSGSALAGVVPWGGGNGHNGGNSGPDSTLSIYQYGSANAALYDQ 60
DB 1 MKLLKVAAPAAIVVSGSALAGVVPWGGGNGHNGGNSGPDSTLSIYQYGSANAALYDQ 60
QY 61 LVTRVVTHEMAHAGYNGADVGQGDADNSTIETQTQNGFRNNATIDQNAKNSDITVGYGG 120
DB 61 SDARKSETTITQSGYNGADVGQGDADNSTIETQTQNGFRNNATIDQNAKNSDITVGYGG 120
QY 121 NNAALVNOTASDSSVMVROVGFNNATANQY 151
DB 121 NNAALVNOTASDSSVMVROVGFNNATANQY 151

RESULT 13

AAB36348
ID AAB36348 standard; protein; 151 AA.
XX AAB36348;
XX 26-FEB-2001 (first entry)
XX AgfA::PT3#3 amino acid sequence SEQ ID NO:16.
XX Salmonella; agfA; chromosomal gene replacement; fimbria; epitope;
KW vaccine; immune response; immunogen.
XX Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX WO2000060102-A2.
XX 12-OCT-2000.
XX 05-APR-2000; 2000NO-CA000356.
XX 05-APR-1999; 99US-0127888P.
XX (UYVI-) UNIV VICTORIA.
XX White AP, Doran JU, Collison SK, Kay WW;
XX WFI; 2000-672631/65.

DR N-PSDB; AAC64624.
XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX Disclosure; Page 136; 139pp; English.

XX The present invention describes a recombinant agfA gene (1) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention

XX Sequence 151 AA;

Query Match 73.8%; Score 577; DB 3; Length 151;
Best Local Similarity 80.8%; Pred. No. 1.3e-47;
Matches 122; Conservative 5; Mismatches 24; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPWGGGNGHNGGNSGPDSTLSIYQYGSANAALYDQ 60
DB 1 MKLLKVAAPAAIVVSGSALAGVVPWGGGNGHNGGNSGPDSTLSIYQYGSANAALYDQ 60
QY 61 LVTRVVTHEMAHAGYNGADVGQGDADNSTIETQTQNGFRNNATIDQNAKNSDITVGYGG 120
DB 61 SDARKSETTITQSGYNGADVGQGDADNSTIETQTQNGFRNNATIDQNAKNSDITVGYGG 120
QY 121 NNAALVNOTASDSSVMVROVGFNNATANQY 151
DB 121 NNAALVNOTASDSSVMVROVGFNNATANQY 151

RESULT 14

AAB36343
ID AAB36343 standard; protein; 151 AA.
XX AAB36343;
XX 26-FEB-2001 (first entry)
XX Escherichia coli CsgA amino acid sequence SEQ ID NO:7.
XX Salmonella; agfA; chromosomal gene replacement; fimbria; epitope;
KW vaccine; immune response; immunogen.
XX Escherichia coli.
XX WO2000060102-A2.
XX 12-OCT-2000.

PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888F.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay WW;
 XX
 DR WPI; 2000-672631/65.
 DR N-PSDB; AAC64619.
 XX
 XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant AgfA
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 135; 139pp; English.
 XX
 CC The present invention describes a recombinant agfA gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEf17/TAf) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant AgfA
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;
 Query Match 65.9%; Score 523; DB 3; Length 151;
 Best Local Similarity 68.9%; Pred. No. 2e-42;
 Matches 104; Conservative 18; Mismatches 29; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHGGNSGSPDSTLSIYQYGSANAALYDQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQYGGGNGHGGNSGSPNSSEINLYYGGNSALALQ 60
 QY 61 LVTRVVTHEMAHAGYNGADVCGQADNSTIELTQNGFRNNTATIDOWNAKNSDITVQYGG 120
 DB 61 TDARNSDLTITQHGGGNGADVCGQSDSSIDLTRQFGNSATLDQWNGKNSMTYKQFGG 120
 QY 121 NNAALVNQATSDSSVMVROVGFNNATANQY 151
 DB 121 GNGRAVDQATSNSSVNVTVQVGFNNATAHQY 151
 RESULT 15
 ABR82651
 ID ABR82651 standard; protein; 151 AA.
 XX
 AC ABR82651;
 XX
 DT 04-DEC-2003 (first entry)
 XX
 DE E. coli CsgA subunit 15 kDa protein.
 XX

KW Plasma protein; immune response; antibacterial; vaccine; gene therapy.
 XX
 OS Escherichia coli.
 XX
 PN WO2003064446-A2.
 XX
 PD 07-AUG-2003.
 XX
 XX 30-JAN-2003; 2003WO-BP000943.
 XX
 XX 31-JAN-2002; 2002GB-00002275.
 PR
 PA (HANS-) HANSA MEDICAL RES AB.
 XX
 PI Bjoerck L, Olsen A, Wikstroem M, Herwald H;
 XX
 DR WPI; 2003-646136/61.
 DR N-PSDB; ACF36153.
 XX
 XX New isolated peptide capable of binding a mammalian plasma protein,
 PT useful in the manufacture of a medicament for the prevention and/or
 PT treatment of a bacterial infection, such as Escherichia coli, Salmonella
 PT or Shigella infections.
 XX
 PS Disclosure; Page 41-42; 42pp; English.
 XX
 CC The invention relates to an isolated peptide capable of binding a
 CC mammalian plasma protein or of generating an immune response in a mammal
 CC selected from sequences shown in ABR82642, ABR82648-49. The peptide or
 CC antibody is useful for treating a bacterial infection in a human or
 CC animal or in the manufacture of a medicament for the prophylactic
 CC treatment of a bacterial infection, such as Escherichia coli, Salmonella
 CC or Shigella infection. The peptide that is immobilized on a solid support
 CC is also useful as a reagent for determining the ability of a plasma
 CC protein to bind to bacteria. The present sequence represents an E. coli
 CC 15 kDa protein
 XX
 SQ Sequence 151 AA;
 Query Match 66.2%; Score 518; DB 7; Length 151;
 Best Local Similarity 68.2%; Pred. NO. 6.2e-42;
 Matches 103; Conservative 18; Mismatches 30; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHGGNSGSPDSTLSIYQYGSANAALYDQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQYGGGNGHGGNSGSPNSSEINLYYGGNSALALQ 60
 QY 61 LVTRVVTHEMAHAGYNGADVCGQADNSTIELTQNGFRNNTATIDOWNAKNSDITVQYGG 120
 DB 61 TDARNSDLTITQHGGGNGADVCGQSDSSIDLTRQFGNSATLDQWNGKNSMTYKQFGG 120
 QY 121 NNAALVNQATSDSSVMVROVGFNNATANQY 151
 DB 121 GNGRAVDQATSNSSVNVTVQVGFNNATAHQY 151
 Search completed: August 2, 2004, 14:48:28
 Job time : 45.9 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:40:48 ; Search time 12 Seconds
(without alignments)
649.627 Million cell updates/sec

Title: US-09-543-407-26

Perfect score: 782

Sequence: 1 MKLLKVAFAAIIVSGSALA.....DSSVMVRQVGFNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/ptodata/2/iaa/5A-COMB.pep.*

2: /cgn2_6/ptodata/2/iaa/5B-COMB.pep.*

3: /cgn2_6/ptodata/2/iaa/6A-COMB.pep.*

4: /cgn2_6/ptodata/2/iaa/6B-COMB.pep.*

5: /cgn2_6/ptodata/2/iaa/PCTUS-COMB.pep.*

6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	687	87.9	151	1	US-08-233-788A-59
2	507	64.8	120	1	US-08-233-788A-57
3	100	12.8	943	4	US-09-036-556-204
4	100	12.8	943	4	US-09-072-596-199
5	100	12.8	943	4	US-09-477-135A-131
6	100	12.8	943	4	US-09-072-967-204
7	98.5	12.6	738	3	US-08-864-038A-3
8	93	11.9	892	4	US-09-336-447A-5
9	91.5	11.7	975	4	US-09-328-352-4764
10	88	11.3	273	4	US-09-328-352-6167
11	86	11.0	1415	4	US-09-252-931A-26438
12	82	10.5	673	3	US-09-196-387-8
13	82	10.5	673	4	US-09-841-835-8
14	82	10.5	949	3	US-09-196-387-10
15	82	10.5	949	4	US-09-841-835-10
16	82	10.5	1327	3	US-09-196-387-2
17	82	10.5	1327	4	US-09-841-835-2
18	82	10.5	1327	4	US-09-972-115A-8
19	81.5	10.4	873	4	US-09-336-447A-13
20	81	10.4	2736	4	US-09-252-931A-30227
21	80	10.2	863	4	US-09-489-039A-10179
22	79	10.1	353	2	US-08-687-702-37
23	78.5	10.0	745	4	US-09-336-115C-6
24	78.5	10.0	1739	4	US-09-540-236-3739
25	78	10.0	1216	4	US-09-134-000C-5130
26	77.5	9.9	321	4	US-09-498-520A-18
27	77.5	9.9	713	4	US-09-059-584-53

28	77	9.8	1338	2	US-08-728-470-9	Sequence 9, Appli
29	77	9.8	1338	3	US-08-719-641-9	Sequence 9, Appli
30	77	9.8	1529	2	US-08-728-470-10	Sequence 10, Appl
31	77	9.8	1529	3	US-08-719-641-10	Sequence 10, Appl
32	77	9.8	1599	2	US-08-617-697-9	Sequence 9, Appli
33	77	9.8	1600	2	US-08-617-697-10	Sequence 10, Appl
34	77	9.8	2315	4	US-08-543-631A-5434	Sequence 5434, Ap
35	76.5	9.8	363	1	US-08-458-023B-6	Sequence 6, Appli
36	76.5	9.8	435	2	US-08-331-515A-2	Sequence 2, Appli
37	76.5	9.8	435	3	US-09-168-406A-2	Sequence 2, Appli
38	76.5	9.8	941	4	US-09-336-447A-9	Sequence 9, Appli
39	76.5	9.8	1612	1	US-08-169-927-2	Sequence 2, Appli
40	76	9.7	364	1	US-07-792-259-17	Sequence 17, Appl
41	76	9.7	1385	1	US-07-876-280-2	Sequence 2, Appli
42	76	9.7	1385	1	US-07-675-772-2	Sequence 2, Appli
43	76	9.7	1385	1	US-08-063-170-2	Sequence 2, Appli
44	76	9.7	1385	1	US-08-158-232-2	Sequence 2, Appli
45	76	9.7	1385	1	US-08-304-626-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-233-788A-59
; Sequence 59, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, Karen S.
; APPLICANT: Clouchier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; TITLE OF INVENTION: OF SALMONELLA
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,788A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEDDANBERRY
; INFORMATION FOR SEQ ID NO: 59:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 151 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-233-788A-59

Query Match 87.9%; Score 687; DB 1; Length 151;
Best Local Similarity 90.1%; Pred. No. 78-60;
Matches 136; Conservative 2; Mismatches 13; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIIVSGSALAGVVPQGGGNNHGGNSGPDSTLSIYQYGSANAALYDQ 60
|||||

Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHGGSSGPDSTLSIYQGSANAALALQ 60
 QY 61 LVTRVVVTHEMAHAGYNGADVQGGADNSTIETQNGFRNATIDOWNAKNSDITVGOYGG 120
 Db 61 SPARKSETTITSGYNGADVQGGADNSTIETQNGFRNATIDOWNAKNSDITVGOYGG 120
 QY 121 NNAALVNOTASDSSVNVQVGGNNATANQY 151
 Db 121 NNPALVNOTASDSSVNVQVGGNNATANQY 151

RESULT 2

US-08-233-788A-57
 ; Sequence 57, Application US/08233788A
 ; Patent No. 5635617
 ; GENERAL INFORMATION:
 ; APPLICANT: Doran, James L.
 ; APPLICANT: Kay, William W.
 ; APPLICANT: Collinson, Karen S.
 ; APPLICANT: Clouchier, Sharon C.
 ; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
 ; OF SALMONELLA
 ; NUMBER OF SEQUENCES: 61
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Seed and Berry
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue
 ; CITY: Seattle
 ; STATE: Washington
 ; COUNTRY: U.S.A.
 ; ZIP: 98104-7092
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/233,788A
 ; FILING DATE: 26-APR-1994
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: King, Joshua
 ; REGISTRATION NUMBER: 35,570
 ; REFERENCE/DOCKET NUMBER: 920043.403C2
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (206) 622-4900
 ; TELEFAX: (206) 682-6031
 ; TELEX: 3723836 SREDANBERRY
 ; INFORMATION FOR SEQ ID NO: 57:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 120 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-233-788A-57

Query Match 64.8%; Score 507; DB 1; Length 120;
 Best Local Similarity 87.5%; Pred. No. 2.1e-42;
 Matches 98; Conservative 2; Mismatches 12; Indels 0; Gaps 0;
 QY 22 VVPQWGGGNGHGGSSGPDSTLSIYQGSANAALYDQVTRVTHEMAHAGYNGADV 81
 Db 1 VVPQWGGGNGHGGSSGPDSTLSIYQGSANAALALQSDARKSETTITQSGYNGADV 60
 QY 82 QGGADNSTIETQNGFRNATIDOWNAKNSDITVGOYGGNNAALVNOTASDS 133
 Db 61 QGGADNSTIETQNGFRNATIDOWNAKNSDITVGOYGGNNAALVNOTASDS 112

RESULT 3

US-09-056-556-204
 ; Sequence 204, Application US/09056556
 ; Patent No. 6350456
 ; GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
 APPLICANT: Skeiky, Yasir A.W.
 APPLICANT: Dillon, Davin C.
 TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
 NUMBER OF SEQUENCES: 241
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: SEED AND BERRY LLP
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington
 COUNTRY: USA
 ZIP: 98104-7092

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/056,556
 FILING DATE: 07-APR-1998
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Maki, David J.
 REGISTRATION NUMBER: 31,392
 REFERENCE/DOCKET NUMBER: 210121.457
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 622-4900
 TELEFAX: (206) 682-6031
 INFORMATION FOR SEQ ID NO: 204:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 943 amino acids
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: linear
 US-09-056-556-204

Query Match 12.8%; Score 100; DB 4; Length 943;
 Best Local Similarity 26.0%; Pred. No. 0.2;
 Matches 40; Conservative 12; Mismatches 52; Indels 50; Gaps 7;
 QY 16 GSALAGVFPQWGGG-GNHN-GCGNSSGPDSTLSIYQGSANAALYDQVTRVTHEMAHA 73
 Db 464 GSCNIGFVNVGSGSLGNNIGSGN-----LGITNIGFNVGDY-----NV 503
 QY 74 GYNGADVQGGADNSTIETQNGFRNATIDOWNAKNSDITVGOYGGN-----121
 Db 504 GFNGAGDFNQGFANT-----GNNNIGFANTGNNNIGIGLSGDNQGGFNIAAGWNS 553
 QY 122 ---NNAALVNOTASDSSVM---VRQVGFNNATAN 149
 Db 554 GTGNSGLFNSGTTNVGIFNAGTGVNGVANSSTGN 587

RESULT 4

US-09-072-596-199
 ; Sequence 199, Application US/09072596
 ; Patent No. 6458366
 ; GENERAL INFORMATION:
 ; APPLICANT: Reed, Steven G.
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Campos-Neto, Antonia
 ; APPLICANT: Houghton, Raymond
 ; APPLICANT: Vedvick, Thomas S.
 ; APPLICANT: Twardzik, Daniel R.
 ; APPLICANT: Lodes, Michael J.
 ; APPLICANT: Hendrickson, Ronald C.
 ; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
 ; TUBERCULOSIS
 ; NUMBER OF SEQUENCES: 350
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SEED AND BERRY LLP
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue
 ; CITY: Seattle

```

; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 199:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 943 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-09-072-596-199

Query Match 12.8%; Score 100; DB 4; Length 943;
Best Local Similarity 26.0%; Pred. No. 0.2;
Matches 40; Conservative 12; Mismatches 52; Indels 50; Gaps 7;

QY 16 GSALAGVVPQWGGG-GNHN-GGNSGSPDSTLSIYQYGSANAALYDQLVTRVVTHEMAHA 73
Db 464 GSGNIGVFNVGSGSLGNYNIGSGN-----LGIYNIGFNGVDY-----NV 503

QY 74 GYGNAGDVQGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGYQYGGN-----121
Db 504 GFGNAGDFNQGFANT-----GNNNIGFANTGNNNIGIGLSGDNQOQGFNIASGWS 553

QY 122 ---NAALVNQTASDSSVM---VRQVGFNNATAN 149
Db 554 GTGNSGLFNSGTNNVGFNAGTGNVGIANSGTGN 587

RESULT 5
US-09-477-135A-131
; Sequence 131, Application US/09477135A
; Patent No. 6572865
; GENERAL INFORMATION:
; APPLICANT: Nano, Francis
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
; TITLE OF INVENTION: Immunostimulatory Peptides
; FILE REFERENCE: 52888
; CURRENT APPLICATION NUMBER: US/09/477,135A
; CURRENT FILING DATE: 2000-01-03
; PRIOR APPLICATION NUMBER: 08990823
; PRIOR FILING DATE: 1997-12-15
; PRIOR APPLICATION NUMBER: US 96/10375
; PRIOR FILING DATE: 1996-06-14
; PRIOR APPLICATION NUMBER: 60/000,254
; PRIOR FILING DATE: 1995-06-15
; NUMBER OF SEQ ID NOS: 169
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 131
; LENGTH: 943
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; US-09-477-135A-131

Query Match 12.8%; Score 100; DB 4; Length 943;
Best Local Similarity 26.0%; Pred. No. 0.2;
Matches 40; Conservative 12; Mismatches 52; Indels 50; Gaps 7;

QY 16 GSALAGVVPQWGGG-GNHN-GGNSGSPDSTLSIYQYGSANAALYDQLVTRVVTHEMAHA 73
Db 464 GSGNIGVFNVGSGSLGNYNIGSGN-----LGIYNIGFNGVDY-----NV 503

QY 74 GYGNAGDVQGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGYQYGGN-----121
Db 504 GFGNAGDFNQGFANT-----GNNNIGFANTGNNNIGIGLSGDNQOQGFNIASGWS 553

QY 122 ---NAALVNQTASDSSVM---VRQVGFNNATAN 149
Db 554 GTGNSGLFNSGTNNVGFNAGTGNVGIANSGTGN 587

RESULT 6
US-09-072-967-204
; Sequence 204, Application US/09072967
; Patent No. 6592877
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Twardzik, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 204:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 943 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-09-072-967-204

Query Match 12.8%; Score 100; DB 4; Length 943;
Best Local Similarity 26.0%; Pred. No. 0.2;
Matches 40; Conservative 12; Mismatches 52; Indels 50; Gaps 7;

QY 16 GSALAGVVPQWGGG-GNHN-GGNSGSPDSTLSIYQYGSANAALYDQLVTRVVTHEMAHA 73
Db 464 GSGNIGVFNVGSGSLGNYNIGSGN-----LGIYNIGFNGVDY-----NV 503

QY 74 GYGNAGDVQGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGYQYGGN-----121
Db 504 GFGNAGDFNQGFANT-----GNNNIGFANTGNNNIGIGLSGDNQOQGFNIASGWS 553

```


Db 393 ASNGEHNHYGIGNGN 407

RESULT 10

US-09-328-352-6167

Sequence 6167, Application US/09328352

Patent No. 6562958

GENERAL INFORMATION:

APPLICANT: Gary L. Breton et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER

TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: GTC99-03PA

CURRENT APPLICATION NUMBER: US/09/328,352

CURRENT FILING DATE: 1999-06-04

NUMBER OF SEQ ID NOS: 8252

SEQ ID NO 6167

LENGTH: 273

TYPE: PRT

ORGANISM: Acinetobacter baumannii

US-09-328-352-6167

Query Match 11.3%; Score 88; DB 4; Length 273;

Best Local Similarity 28.5%; Pred. No. 0.6;

Matches 43; Conservative 19; Mismatches 65; Indels 24; Gaps 7;

Qy 1 MKLLKVA--PAATVSGSALAGVVPWGGGNNH-----NGGNS--GPDSTLSIYQVGS 53

Db 19 MKLAIASALLSALAVSAAANA-YQAEVGGSYNLDPDNGSSVSXKFGVGTGYFNPVQTR 77

Qy 54 NAALYDQVTRVVTHEMAHAGYNGADVGQGDNSTLTQNGFRNNATIDOWNAKNSDI 113

Db 78 NAPLAERAFNLNRASNVNHNHYG-----DMSGTRDTQYGVG-----VEYFVPSDF 123

Qy 114 TVQYGGNNALVNQTSDDSSV--MVRQVGF 142

Db 124 YLSGDVGRNEREIDNTNIDSKVTYYAAEVGY 154

RESULT 11

US-09-252-991A-26438

Sequence 26438, Application US/09252991A

Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A

CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 26438

LENGTH: 1415

TYPE: PRT

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-26438

Query Match 11.0%; Score 86; DB 4; Length 1415;

Best Local Similarity 23.9%; Pred. No. 7.8;

Matches 38; Conservative 21; Mismatches 68; Indels 32; Gaps 5;

Qy 13 VVSGSALAGVVPWGGGNNHNG-----CGNSSGPDSTLSIYQVGSANAALYDQVLT 63

Db 366 ILSHEVSVAAVQANAGDSGVHVAGPAGANSDSNGVTIVQQPFAVDLAAGANGTS 425

Qy 64 RVWTHEMAHAGYV-----NGADVQGDNSTLTQNG-----GFERNATIDQWN 107

Db 426 AVQSGSGANIGSANGISVWQSGANIGAGASDISVVQSQNSNIGSGYNGVTIVQSQN 485

Qy 108 AKN-----SDITVQYGGNNALVNQTSDDSSVMVRQVG 141

Db 486 GANIGSGAGITVWQ--SQNGANIGSGAGISVWQSQSG 522

RESULT 12

US-09-196-387-8

Sequence 8, Application US/09196387

Patent No. 6277613

GENERAL INFORMATION:

APPLICANT: de Lange, Titia

APPLICANT: Smith, Susan

TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS

TITLE OF INVENTION: OF USE THEREOF

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSEE: Klauber & Jackson

STREET: 411 Hackensack Avenue, 4th Floor

CITY: Hackensack

STATE: New Jersey

COUNTRY: USA

ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/196,387

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/095,225

FILING DATE: June 10, 1998

ATTORNEY/AGENT INFORMATION:

NAME: Jackson Esq., David A.

REGISTRATION NUMBER: 26,742

REFERENCE/DOCKET NUMBER: 600-1-230 CIP1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201-487-5800

TELEFAX: 201-343-1684

TELEX: 133521

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 673 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-196-387-8

Query Match 10.5%; Score 82; DB 3; Length 673;

Best Local Similarity 29.6%; Pred. No. 7.4;

Matches 34; Conservative 14; Mismatches 59; Indels 8; Gaps 4;

Qy 6 VAAFAAI-VVSGSALAGVVPWGGGNNHNGGNSGPDSTLSIYQVGSANAALYDQL--- 61

Db 99 VAAAPVPAVSTSAAGVAPNAPAGSGNNPSSSSSPTSS-SSSSPSPGSSLAESPEAA 157

Qy 62 -VTRVVTHEMAHAGYNGADVGQGDNSTLTQNG--FRNNATIDOWNAKNSDI 113

Db 158 GVSSTAPLGGAAGPCTGVPVAVSGALRELLEACRNGDSVRVRLVDAAVNAKDM 212

RESULT 13

US-09-841-835-8

Sequence 8, Application US/09841835

Patent No. 6506587

GENERAL INFORMATION:

APPLICANT: de Lange, Titia

APPLICANT: Smith, Susan

TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS

TITLE OF INVENTION: OF USE THEREOF

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

```
/ ADDRESS: Klauber & Jackson
/ STREET: 411 Hackensack Avenue, 4th Floor
/ CITY: Hackensack
/ STATE: New Jersey
/ COUNTRY: USA
/ ZIP: 07601
/
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent In Release #1.0, Version #1.30
/
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/841.835
/ FILING DATE:
/ CLASSIFICATION:
/
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 09/196.387
/ FILING DATE:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Jackson Esq., David A.
/ REGISTRATION NUMBER: 26,742
/ REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 201-487-5800
/ TELEFAX: 201-343-1684
/ TELEX: 133521
/
/ INFORMATION FOR SEQ ID NO: 8:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 673 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/
/ US-09-841-835-8
/
/ Query Match 10.5%; Score 82; DB 4; Length 673;
/ Best Local Similarity 29.6%; Pred. No. 7.4;
/ Matches 34; Conservative 14; Mismatches 59; Indels 8; Gaps 4;
/
/ QY 6 VAAFAAI-VVSGSALAGVVPQWGGNGNNGSSGPDSTLSIYQVGSANAALYDQI--- 61
/ DB 99 VAAFPVPAVSTSSAAGVAPNPAAGSGNNPSSSSPTSS-SSSSPSPGSSLAESPEAA 157
/
/ QY 62 -VTRVVTHMAHAGYNGADVQGGADNSTIELTQNG--FRNNATIDQWAKNSDI 113
/ DB 158 GVSSTAPLPGNAGPGTGVPVAVSGALRELLACRNGDYSRVKRLVDAANVNAKDM 212
/
/ RESULT 14
/ US-09-196-387-10
/ Sequence 10, Application US/09196387
/ Patent No. 6277613
/ GENERAL INFORMATION:
/ APPLICANT: de Lange, Titia
/ APPLICANT: Smith, Susan
/ TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
/ TITLE OF INVENTION: OF USE THEREOF
/ NUMBER OF SEQUENCES: 12
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Klauber & Jackson
/ STREET: 411 Hackensack Avenue, 4th Floor
/ CITY: Hackensack
/ STATE: New Jersey
/ COUNTRY: USA
/ ZIP: 07601
/
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent In Release #1.0, Version #1.30
/
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/196.387
/ FILING DATE:
/ CLASSIFICATION:
/
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 09/196.387
/ FILING DATE:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Jackson Esq., David A.
/ REGISTRATION NUMBER: 26,742
/ REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 201-487-5800
/ TELEFAX: 201-343-1684
/ TELEX: 133521
/
/ INFORMATION FOR SEQ ID NO: 10:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 949 amino acids
/ TYPE: amino acid
/
/ RESULT 15
/ US-09-841-835-10
/ Sequence 10, Application US/09841835
/ Patent No. 6506587
/ GENERAL INFORMATION:
/ APPLICANT: de Lange, Titia
/ APPLICANT: Smith, Susan
/ TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
/ TITLE OF INVENTION: OF USE THEREOF
/ NUMBER OF SEQUENCES: 12
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Klauber & Jackson
/ STREET: 411 Hackensack Avenue, 4th Floor
/ CITY: Hackensack
/ STATE: New Jersey
/ COUNTRY: USA
/ ZIP: 07601
/
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent In Release #1.0, Version #1.30
/
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/841.835
/ FILING DATE:
/ CLASSIFICATION:
/
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 09/196.387
/ FILING DATE:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Jackson Esq., David A.
/ REGISTRATION NUMBER: 26,742
/ REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 201-487-5800
/ TELEFAX: 201-343-1684
/ TELEX: 133521
/
/ INFORMATION FOR SEQ ID NO: 10:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 949 amino acids
/ TYPE: amino acid
```

```

; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-841-835-10

Query Match      10.5%; Score 82; DB 4; Length 949;
Best Local Similarity 29.6%; Pred. No. 12;
Matches 34; Conservative 14; Mismatches 59; Indels 8; Gaps 4;

QY 6 VAAFAAI-VVSGSALAGVVPQWGGGHHNGGNSGPDSTLSIYQGSANAALYDQL--- 61
   | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 99 VAAAPVVPVAVSTSSAAGVAPNPAGSGGNNPSSSSSPTSS--SSSSPSPGSSLAESPEAA 157

QY 62 -VTRVTVTHVAHAGYGNQADVGQADNSTLTQNG--FRNNATIDOWNAKNSDI 113
   | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 158 GVSSTAPLGPGAAGPGTGVPVAVSGALRELLAEACRNGDVSRVKRVLVDAAVNAKDM 212

```

Search completed: August 2, 2004, 14:58:36
Job time : 13 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:54:48 ; Search time 36.8 Seconds
(without alignments)
1287.123 Million cell updates/sec

Title: US-09-543-407-26
Perfect score: 782
Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVGFNNATANQY 151

Scoring table: BLOSUM62
Gap 10.0 , Gapext 0.5

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	520	66.5	151	12	US-09-741-873B-4
2	520	66.5	151	12	US-09-741-873B-4
3	442	56.5	131	12	US-09-741-873B-2
4	442	56.5	131	12	US-09-741-873B-2
5	116.5	14.9	445	15	US-10-369-493-20638
6	105	13.4	597	9	US-09-793-306-146
7	100	12.8	943	9	US-09-966-834-131
8	100	12.8	943	10	US-09-997-182-131
9	100	12.8	943	10	US-09-997-181-131
10	100	12.8	943	14	US-10-193-002-199
11	100	12.8	943	14	US-10-084-843-204
12	100	12.8	3300	12	US-10-282-122A-64369
13	98.5	12.5	486	15	US-10-369-493-20619
14	96	12.3	186	12	US-10-282-122A-49412
15	93	11.9	892	10	US-09-952-267-5

16	92.5	11.8	253	16	US-10-437-963-114193
17	92.5	11.8	1721	12	US-10-282-122A-62548
18	92.5	11.8	354	10	US-09-820-843A-21
19	92	11.8	616	16	US-10-437-963-193067
20	91.5	11.7	154	16	US-10-437-963-162284
21	91.5	11.7	591	14	US-10-233-553-23
22	91.5	11.7	894	14	US-10-233-553-11
23	91	11.6	678	12	US-10-282-122A-64573
24	91	11.6	974	12	US-10-282-122A-44999
25	91	11.6	1106	12	US-10-282-122A-62472
26	91	11.6	6310	12	US-10-282-122A-67793
27	88.5	11.3	197	12	US-10-425-114-67750
28	88.5	11.3	1621	14	US-10-185-990-10
29	88.5	11.3	1626	14	US-10-185-990-11
30	88	11.3	1448	16	US-10-408-765A-998
31	87.5	11.2	204	12	US-10-424-599-203972
32	87.5	11.2	250	16	US-10-479-670-164
33	87.5	11.2	1649	15	US-10-369-493-18460
34	87	11.1	904	15	US-10-369-493-12420
35	86.5	11.1	2204	12	US-10-282-122A-64364
36	86	11.0	275	12	US-10-424-599-215142
37	85.5	10.9	263	12	US-10-425-114-49960
38	85.5	10.9	278	9	US-09-810-264-28
39	85	10.9	193	16	US-10-437-963-148500
40	84.5	10.8	126	16	US-10-479-670-80
41	84	10.7	191	16	US-10-437-963-105413
42	84	10.7	191	16	US-10-437-963-152921
43	84	10.7	353	16	US-10-437-963-120176
44	84	10.7	391	12	US-10-424-599-148448
45	84	10.7	394	12	US-10-425-114-55337

ALIGNMENTS

RESULT 1

US-09-741-873B-4
; Sequence 4, Application US/09741873B
; Publication No. US20020081722A1
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09741,873B
; PRIOR FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-4

Query Match 66.5%; Score 520; DB 12; Length 151;
Best Local Similarity 68.2%; Pred No. 2,2e-44;
Matches 103; Conservative 19; Mismatches 29; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNSGPDSTLSIYQYGSANAALYDQ 60

Db 1 MKLLKVAALIAVFGSSAVAGVVPQYGGGNGHGGGNSGENSELNIYQYGGNSALALQ 60
QY 61 LVTRVVTTHMAHAGVNGADVCGQADNSTIELTQNGFRNNATIDQNAKNSDITVQYGG 120
Db 61 TDARNSDLTITQHGNGGADVCGQSDSSIDLQRFNGSATLDQNGKNSMTVQKQFGG 120
QY 121 NNAALVNQTSADSSVMVQVGFNNATANQY 151
Db 121 GNGRAVDQTSANSSVNVTVQVGFNNATAHQY 151

RESULT 2
US-09-741-873B-4
; Sequence 4, Application US/09741873B
; Publication No. US20040096965A9
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; CURRENT APPLICATION NUMBER: US/09/741,873B
; CURRENT FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-4

Query Match 66.5%; Score 520; DB 12; Length 151;
Best Local Similarity 68.2%; Pred. No. 2.2e-44;
Matches 103; Conservative 19; Mismatches 29; Indels 0; Gaps 0;
QY 1 MKLLKVAALIAVFGSSAVAGVVPQYGGGNGHGGGNSGENSELNIYQYGSANAALYDQ 60
Db 1 MKLLKVAALIAVFGSSAVAGVVPQYGGGNGHGGGNSGENSELNIYQYGGNSALALQ 60
QY 61 LVTRVVTTHMAHAGVNGADVCGQADNSTIELTQNGFRNNATIDQNAKNSDITVQYGG 120
Db 61 TDARNSDLTITQHGNGGADVCGQSDSSIDLQRFNGSATLDQNGKNSMTVQKQFGG 120
QY 121 NNAALVNQTSADSSVMVQVGFNNATANQY 151
Db 121 GNGRAVDQTSANSSVNVTVQVGFNNATAHQY 151

RESULT 3
US-09-741-873B-2
; Sequence 2, Application US/09741873B
; Publication No. US20020061722A1
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; CURRENT FILING DATE: 2003-04-04

; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-2
Query Match 56.5%; Score 442; DB 12; Length 131;
Best Local Similarity 64.9%; Pred. No. 1.3e-36;
Matches 85; Conservative 18; Mismatches 28; Indels 0; Gaps 0;
QY 21 GVPQWGGGNGHGGGNSGPDSTLSIYQYGSANAALYDQLVTRVVTTHMAHAGVNGAD 80
Db 1 GVPQYGGGNGHGGGNSGENSELNIYQYGGNSALALQTDARNSDLTITQHGNGGAD 60
QY 81 VQGDADNSTIELTQNGFRNNATIDQNAKNSDITVQYGGNNAALVNQTSADSSVMVQV 140
Db 61 VQGSDDSSIDLQRFNGSATLDQNGKNSMTVQYGGGNGAAVDQTSANSSVNVTVQV 120
QY 141 GFGNNATANQY 151
Db 121 GFGNNATAHQY 131

RESULT 4
US-09-741-873B-2
; Sequence 2, Application US/09741873B
; Publication No. US20040096965A9
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; CURRENT FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-2
Query Match 56.5%; Score 442; DB 12; Length 131;

Best Local Similarity	64.9%;	Pred. No.	1.3e-35;
Matches	85;	Conservative	18; Mismatches 28; Indels 0; Gaps 0;
QY	21	GVPQWGGGNGHGGNSGSDLSISIQYGSANAALVDLTVRVVTHETAGACYNCAAD	80
		: : : : : :	
Dd	1	GVPQYVGCGNHGGGNNSGNSELTNYQYGGGNSALACTDARNSDLIITQHGGGNCAD	60
		: : : : : :	
QY	81	VGGGADNSTIELTONGFENNATIDOWAKNSDITVGYQGNNAAALVNQTASDSVVNVTVQV	140
		: : : : : :	
Dd	61	VCGGDDSDSIDLTRGFNGSATLDQWGNKSEMTVKQFGGGGAAYDVDTASNSSVNVTVQV	120
		: : : : : :	
QY	141	GFGNNTANQY	151
		: : : : : :	
Dd	121	GFGNNTAFAQY	131
		: : : : : :	

```

RESULT 5
US-10-369-493-20638
; Sequence 20638, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 20638
; LENGTH: 445
; TYPE: PRT
; ORGANISM: Rhodospseudomonas palustris
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(445)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-20638

```

```

Query Match      14.9%; Score 116.5; DB 15; Length 445;
Best Local Similarity 27.5%; Pred.No. 0.0032;
Matches 46; Conservative 20; Mismatches 60; Indels 41; Gaps 6

QY      7 AAFRA-----LVVGSALAGVWPQWGGG-----NHNGG-----GNSGPDSTLSIYY 50
      ||||| : : : : :
Db      19 AAFRADSTVYLNTQDNCQANITQSGNGSVGAFNGSGFLQENGTLSCA-NLLTVKGS 77
      ||||| : : : : :

QY      51 GSANAALYDQLVTRYVTHMAFAGYGGADVGQAGDADNSTIELTQNGFRNNATIDQNNAK 110
      ||| : : : : :
Db      78 GNSNSVGRD-----IQGKSGAGNSAAIFQEGTGTSDVELQQTGTSCNCAVPSGHNWTN 129
      ||| : : : : :

QY      111 -----SDITVGQYGGNNAALVNQATSDSNVVRQVG 141
      : : : : :
Db      130 DEGVENKTIQDSSNNGSKSVYIQGKKNVFSIKQKNTGNTSVNOIG 176
      : : : : :

```

```

RESULT 6
US-09-793-306-146
; Sequence 146, Application US/097933306
; Patent No. US20020098200A1
; GENERAL INFORMATION:
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Skeiky, Yasir
; APPLICANT: Ovendale, Pamela
; APPLICANT: Jen, Shyian
; APPLICANT: Lodes, Michael
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compounds and Methods for Diagnosis and Immunotherapy

```

```

? TITLE OF INVENTION: of Tuberculosis
? FILE REFERENCE: 014058-008740US
? CURRENT APPLICATION NUMBER: US/09/793,306
? CURRENT FILING DATE: 2001-02-26
? PRIOR APPLICATION NUMBER: US 60/185,037
? PRIOR FILING DATE: 2000-02-25
? PRIOR APPLICATION NUMBER: US 60/223,828
? PRIOR FILING DATE: 2000-08-08
? NUMBER OF SEQ ID NOS: 164
? SOFTWARE: PatentIn Ver. 2.1
? SEQ ID NO 146
? LENGTH: 597
? TYPE: PRT
? ORGANISM: Artificial Sequence
? FEATURE:
? OTHER INFORMATION: Description of Artificial Sequence: mITC#3-His
US-09-793-306-146

```

```

Query Match      13.4%; Score 105; DB 9; Length 597;
Best Local Similarity 29.0%; Pred. No. 0.065;
Matches 40; Conservative 18; Mismatches 58; Indels 22; Gaps 7

QY 14 VSGSALAGVWQGGGNNHRCGSSPFDSTLSIYQGSANAALYDQLVTRVVTHEMAHA 73
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 338 LTGMDLVIGALNSGIGN-LGFGNSG--NNNIGFNSGNNNVGFNS-----GNNNF 386
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 74 GYGNAGADVQGAADNSTIELTONGFNNAATIDQ--WNAKSDITVGYQGNNAALVNAQTAS 131
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 387 GFGNAGDINTGFGNAGD--INTGFGNAGFGNMGIGNAGNEDMGVGGSGFNVGVN--AG 442
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 132 DSSVMVROVFGGNATAN 149
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 443 NOS-----VGFNGAGTLN 455
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 7
US-09-996-634-131
; Sequence 131, Application US/09956634
; Patent No. US20020172684A1
; GENERAL INFORMATION:
; APPLICANT: Nano, Francis
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
; TITLE OF INVENTION: Immunostimulatory Peptides
; FILE REFERENCE: 61260
; CURRENT APPLICATION NUMBER: US/09/996,634
; CURRENT FILING DATE: 2001-11-28
; PRIOR APPLICATION NUMBER: 09/447,135
; PRIOR FILING DATE: 2000-01-03
; PRIOR APPLICATION NUMBER: 08/990,823
; PRIOR FILING DATE: 1997-12-15
; PRIOR APPLICATION NUMBER: US 96/10375
; PRIOR FILING DATE: 1996-06-14
; PRIOR APPLICATION NUMBER: 60/000,254
; PRIOR FILING DATE: 1995-06-15
; NUMBER OF SEQ ID NOS: 169
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 131
; LENGTH: 943
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-996-634-131

```

	Query Match	12.8%	Score 100;	DB 9;	Length 943;
	Best Local Similarity	26.0%	Pred. No. 0.37;		
	Matches	40;	Conservative	12;	Mismatches 52; Indels 50; Gaps 7;
QY	16	GSALAGVVPQWGG-GNNH-GGNSGSPDSTLSIVYGSANAALVDQIVTRVVTHEMAHA	73		
Db	571	GSGNIGFNVGSGSLGNYNGSN-----LGIYNIQFNGVGDY-----	NV	610	
QY	74	QYGNAGDVGGGANDSTIELTQNGFRNNALIDWNAKNSDITVQVGGN-----	121		
Db	611	GFGNAGDFNQGFAFNT-----GNNNIGFANTGNNIGIGISGDRGCGFNIAAGWNS	660		

Query Match 12.8%; Score 100; DB 14; Length 943;
Best Local Similarity 26.0%; Pred. No. 0.37;
Matches 40; Conservative 12; Mismatches 52; Indels 50; Gaps 7;

QY 16 GSALAGVVPQGGG-GNHN-GGGNSSGPDSTLSIYQGSANAALYDQLVTRVVTHEMAHA 73
DB 464 GSGNIGVFNVGSGSLGNYNIGSGN-----LGYNIGFNGVDY-----NV 503

QY 74 GYGNGADVQGGADNSTDITLQNGFRNNATIDQWNKNSDITVGOYGGN-----121
DB 504 GFGNAGDFNQGFANT-----GNNIGFANTGNNNIGLIGLSDNQOGENIASGWNS 553

QY 122 ---NAALVNQTASDSSVM---VRQVGFNNATAN 149
DB 554 GTGNSGLFNSGTNNVGIFNAGTGNVGIANS GTGN 587

RESULT 11
US-10-084-843-204
; Sequence 204, Application US/10084843
; Publication No. US20030143243A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; Skeiky, Yasir A.W.
; Dillon, Davin C.
; Campos-Neto, Antonio
; Houghton, Raymond
; Vedvick, Thomas S.
; Twardzik, Daniel R.
; Lodes, Michael J.
; Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESSES:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
City: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/084,843
FILING DATE: 25-Feb-2002
CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/072,967
FILING DATE: 05-MAY-1998
; ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 204:
SEQUENCE CHARACTERISTICS:
LENGTH: 943 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 204:

US-10-084-843-204

Query Match 12.8%; Score 100; DB 14; Length 943;
Best Local Similarity 26.0%; Pred. No. 0.37;
Matches 40; Conservative 12; Mismatches 52; Indels 50; Gaps 7;

QY 16 GSALAGVVPQGGG-GNHN-GGGNSSGPDSTLSIYQGSANAALYDQLVTRVVTHEMAHA 73
DB 464 GSGNIGVFNVGSGSLGNYNIGSGN-----LGYNIGFNGVDY-----NV 503

QY 74 GYGNGADVQGGADNSTDITLQNGFRNNATIDQWNKNSDITVGOYGGN-----121
DB 504 GFGNAGDFNQGFANT-----GNNIGFANTGNNNIGLIGLSDNQOGENIASGWNS 553

QY 122 ---NAALVNQTASDSSVM---VRQVGFNNATAN 149
DB 554 GTGNSGLFNSGTNNVGIFNAGTGNVGIANS GTGN 587

RESULT 12
US-10-282-122A-64369
; Sequence 64369, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Chlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See file Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 64369
; LENGTH: 3300
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-10-282-122A-64369

Query Match 12.8%; Score 100; DB 12; Length 3300;
Best Local Similarity 26.0%; Pred. No. 1.7;
Matches 40; Conservative 12; Mismatches 52; Indels 50; Gaps 7;

QY 16 GSALAGVVPQGGG-GNHN-GGGNSSGPDSTLSIYQGSANAALYDQLVTRVVTHEMAHA 73
DB 590 GSGNIGVFNVGSGSLGNYNIGSGN-----LGYNIGFNGVDY-----NV 629

QY 74 GYGNGADVQGGADNSTDITLQNGFRNNATIDQWNKNSDITVGOYGGN-----121

```
Db 630 GFGNAGDENQCFANT-----GNNIGFANTGNNIGLIGSGDNQGFNIASGWSN 679
Qy 122 ---NAALVNOTASDSSVM---VRQVGFNNATAN 149
Db 680 GTGNSGLFSGTNNVGNFAGTGNVGIANSCTGN 713

RESULT 13
US-10-369-493-20619
; Sequence 20619, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 20619
; LENGTH: 486
; TYPE: PRT
; ORGANISM: Rhodospseudomonas palustris
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(486)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-20619

Query Match 12.6%; Score 98.5; DB 15; Length 486;
Best Local Similarity 25.7%; Pred. No. 0.23;
Matches 37; Conservative 22; Mismatches 50; Indels 35; Gaps 6;

Qy 11 AIVVSGS---ALAGVVPQWGGGNGHNGGNSGP-----DSTLSIYQYGSANAAL 57
Db 86 SVVAGTDYRIVAGVLPGLGVKKAYAGDGSPPFPKDLASDTFFSAPFGADGF- 144
Qy 58 YDQLVTRVVTTHMAHAGVNGADVGQGDADNSTIELTONGFRNNATIDQWNAKNSDITVG- 116
Db 145 -----VILHTAAVGGPRGALISQG--NLLI-----AQSSLVDAMRLTEADVNLGM 188
Qy 117 ---QYGGNNAALVNOTASDSSVM 136
Db 189 LPLFHTVGLGLMLTLQQAGGASVI 212

RESULT 14
US-10-282-122A-49412
; Sequence 49412, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
```

```
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 49412
; LENGTH: 186
; TYPE: PRT
; ORGANISM: Burkholderia fungorum
US-10-282-122A-49412

Query Match 12.3%; Score 96; DB 12; Length 186;
Best Local Similarity 29.3%; Pred. No. 0.12;
Matches 49; Conservative 18; Mismatches 62; Indels 38; Gaps 10;

Qy 1 MKLVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNS-----GPDSTLSIYQYGSANA 56
Db 1 MKL-----SLPAIAAASSLGLVPL-----GAHAADGTISITGTVDSTCSI--NGNANGT 50
Qy 57 LYDQLVTRVVTTHMAHAGVNGADVG--QGADNST-IELTONG-----FRNNAT 102
Db 51 PADKAIT-----LATVPAGSLASAGAVAGTSNPTDLQLSLTGCTGTATKAIAFPENGPT 104
Qy 103 IDQWNAKNSDITVGQYGGNNAALVNOTASDSSVMVRQVGFNNATAN 149
Db 105 VDQTNGYLSN-TAGTAQNVVEVRLNNAQNPINV---TTGANDITN 147

RESULT 15
US-09-952-267-5
; Sequence 5, Application US/09952267
; Publication No. US2003003272A1
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USP41 AND USP42 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/09/952,267
; CURRENT FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: 09/336,447
; PRIOR FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 892
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-952-267-5

Query Match 11.9%; Score 93; DB 10; Length 892;
Best Local Similarity 27.7%; Pred. No. 1.7;
```

	Matches	41;	Conservative	14;	Mismatches	49;	Indels	44;	Gaps	9;
QY	28	GGGNHN	-----	GGGNSS	---GPDSTLSIYQYGSANAALYDQLVTRVVTHEMAHAGY	75				
Db	89	GGDYNEAKGNYS	TVGGSSNTAKGKSTIGGGDTNDANG	-----	TVSTIGGGY	137				
QY	76	-----	GNGADVGGADNSTI	--ELTQGFNNATIDQWNAKUSDITVGOYG	---GNNAL	125				
Db	138	YSRAIGDSSTIGGGY	YNQATGKSTVAGGRNN	----	QATGNNSTVAGGSYNOATGNNSTV	193				
QY	126	V----	NQTASDSSVMVRQVGFNNATAN	149						
Db	194	AGGSHNQATGEGSF	---AAGVENKANAN	218						

Search completed: August 2, 2004, 15:36:13
 Job time : 37.8 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:48:33 ; Search time 167.9 Seconds
(without alignments)
877.809 Million cell updates/sec

Title: US-09-543-407-26

Perfect score: 782
Sequence: 1 MKLLKVAFAAIVSGSALA.....DSSVMVQVGFNNATANCY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 6019581 seqs, 976053577 residues

Total number of hits satisfying chosen parameters: 6019581

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents AA Main.*
1: /cgn2_6/ptodata/2/paa/PCTUS COMB.pcp.*
2: /cgn2_6/ptodata/2/paa/US06 COMB.pcp.*
3: /cgn2_6/ptodata/2/paa/US07 COMB.pcp.*
4: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
5: /cgn2_6/ptodata/2/paa/US081 COMB.pcp.*
6: /cgn2_6/ptodata/2/paa/US082 COMB.pcp.*
7: /cgn2_6/ptodata/2/paa/US083 COMB.pcp.*
8: /cgn2_6/ptodata/2/paa/US084 COMB.pcp.*
9: /cgn2_6/ptodata/2/paa/US085 COMB.pcp.*
10: /cgn2_6/ptodata/2/paa/US086 COMB.pcp.*
11: /cgn2_6/ptodata/2/paa/US087 COMB.pcp.*
12: /cgn2_6/ptodata/2/paa/US088 COMB.pcp.*
13: /cgn2_6/ptodata/2/paa/US089 COMB.pcp.*
14: /cgn2_6/ptodata/2/paa/US090 COMB.pcp.*
15: /cgn2_6/ptodata/2/paa/US091 COMB.pcp.*
16: /cgn2_6/ptodata/2/paa/US092 COMB.pcp.*
17: /cgn2_6/ptodata/2/paa/US093 COMB.pcp.*
18: /cgn2_6/ptodata/2/paa/US094 COMB.pcp.*
19: /cgn2_6/ptodata/2/paa/US095 COMB.pcp.*
20: /cgn2_6/ptodata/2/paa/US096 COMB.pcp.*
21: /cgn2_6/ptodata/2/paa/US097A COMB.pcp.*
22: /cgn2_6/ptodata/2/paa/US097B COMB.pcp.*
23: /cgn2_6/ptodata/2/paa/US098 COMB.pcp.*
24: /cgn2_6/ptodata/2/paa/US099A COMB.pcp.*
25: /cgn2_6/ptodata/2/paa/US099B COMB.pcp.*
26: /cgn2_6/ptodata/2/paa/US100 COMB.pcp.*
27: /cgn2_6/ptodata/2/paa/US101 COMB.pcp.*
28: /cgn2_6/ptodata/2/paa/US102 COMB.pcp.*
29: /cgn2_6/ptodata/2/paa/US103 COMB.pcp.*
30: /cgn2_6/ptodata/2/paa/US104 COMB.pcp.*
31: /cgn2_6/ptodata/2/paa/US106 COMB.pcp.*
32: /cgn2_6/ptodata/2/paa/US107 COMB.pcp.*
33: /cgn2_6/ptodata/2/paa/US60 COMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description

1	782	100.0	151	19	US-09-543-407-26	Sequence 26, Appl
2	709	90.7	151	19	US-09-543-407-20	Sequence 20, Appl
3	692	88.5	151	19	US-09-543-407-5	Sequence 5, Appl
4	687	87.9	151	6	US-08-233-642A-57	Sequence 57, Appl
5	675	86.3	151	19	US-09-543-407-18	Sequence 18, Appl
6	617	78.9	151	19	US-09-543-407-28	Sequence 28, Appl
7	614	78.5	151	19	US-09-543-407-12	Sequence 12, Appl
8	612	78.3	151	19	US-09-543-407-14	Sequence 14, Appl
9	609	77.9	151	19	US-09-543-407-24	Sequence 24, Appl
10	605	77.4	131	19	US-09-543-407-31	Sequence 31, Appl
11	601	76.9	151	19	US-09-543-407-22	Sequence 22, Appl
12	600	76.7	151	19	US-09-543-407-30	Sequence 30, Appl
13	577	73.8	151	19	US-09-543-407-16	Sequence 16, Appl
14	523	66.9	151	19	US-09-543-407-7	Sequence 7, Appl
15	520	66.5	151	13	US-08-978-878-4	Sequence 4, Appl
16	520	66.5	151	21	US-09-741-873B-4	Sequence 2, Appl
17	518	66.2	151	33	US-60-352-946-2	Sequence 2, Appl
18	518	66.2	151	33	US-60-444-371-2	Sequence 2, Appl
19	507	64.8	120	6	US-08-233-642A-55	Sequence 55, Appl
20	473	60.5	109	19	US-09-543-407-34	Sequence 34, Appl
21	466	59.6	158	16	US-08-252-691-5834	Sequence 5834, Ap
22	466	59.6	158	16	US-03-252-691C-5834	Sequence 5834, Ap
23	466	59.6	158	30	US-10-417-886-5834	Sequence 5834, Ap
24	442	56.5	131	13	US-08-978-878-2	Sequence 2, Appl
25	442	56.5	131	21	US-09-741-873B-2	Sequence 35, Appl
26	338	43.2	109	19	US-09-543-407-35	Sequence 37, Appl
27	266	34.0	68	19	US-09-543-407-37	Sequence 32, Appl
28	201.5	25.8	70	19	US-09-543-407-32	Sequence 30, Appl
29	156	19.9	48	19	US-09-543-407-39	Sequence 30, Appl
30	116.5	14.9	186	16	US-09-252-691-5833	Sequence 5833, Ap
31	116.5	14.9	186	16	US-09-252-691C-5833	Sequence 5833, Ap
32	116.5	14.9	186	30	US-10-417-886-5833	Sequence 5833, Ap
33	116.5	14.9	445	29	US-10-369-493-20638	Sequence 20638, A
34	116.5	14.9	445	33	US-60-360-039-20638	Sequence 20638, A
35	114.5	14.6	145	21	US-09-739-449-8854	Sequence 8854, Ap
36	114.5	14.6	145	23	US-09-803-110-8854	Sequence 8854, Ap
37	112	14.3	151	19	US-09-543-407-8	Sequence 8, Appl
38	107	13.7	151	19	US-09-543-407-6	Sequence 6, Appl
39	105	13.4	597	1	PCT-US01-05992-146	Sequence 146, App
40	105	13.4	597	22	US-09-793-306-146	Sequence 146, App
41	100	12.8	943	1	PCT-US99-03265-199	Sequence 199, App
42	100	12.8	943	1	PCT-US99-03268-204	Sequence 204, App
43	100	12.8	943	13	US-08-942-341-199	Sequence 199, App
44	100	12.8	943	13	US-08-942-578-204	Sequence 204, App
45	100	12.8	943	14	US-09-024-753-199	Sequence 199, App

ALIGNMENTS

RESULT 1
US-09-543-407-26
; Sequence 26, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.

US-09-543-407-26

Query Match 100.0%; Score 782; DB 19; Length 151;
Best Local Similarity 100.0%; Pred. No. 1.8e-75;
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALYDQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALYDQ 60

QY 61 LVTRVVTHEMAHAGYNGADVCGADNSTIELTONGFRNNATIDQWNAKNSDITVQYGG 120
DB 61 LVTRVVTHEMAHAGYNGADVCGADNSTIELTONGFRNNATIDQWNAKNSDITVQYGG 120

QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 2

US-09-543-407-20
Sequence 20, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 20
LENGTH: 151
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.

Query Match 90.7%; Score 709; DB 19; Length 151;
Best Local Similarity 89.9%; Pred. No. 1.3e-67;
Matches 143; Conservative 0; Mismatches 0; Indels 16; Gaps 2;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAAL--- 57
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALQ 60

QY 58 -----YDOLVTRVVTHEMAHAGYNGADVCGADNSTIELTONGFRNNATIDQWNAKNSD 112
DB 61 SDARKYDOLVTRVVTHEMAH-----GGADNSTIELTONGFRNNATIDQWNAKNSD 112

QY 113 ITVGQYGNNAALVNQTASDSSVMVRQVGFNNATANQY 151
DB 113 ITVGQYGNNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 3

US-09-543-407-5
Sequence 5, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407

CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5
LENGTH: 151
TYPE: PRT
ORGANISM: Salmonella enteritidis
US-09-543-407-5

Query Match 88.5%; Score 692; DB 19; Length 151;
Best Local Similarity 90.7%; Pred. No. 8.7e-66;
Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALYDQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALQ 60

QY 61 LVTRVVTHEMAHAGYNGADVCGADNSTIELTONGFRNNATIDQWNAKNSDITVQYGG 120
DB 61 SDARKSETTITQSGYNGADVCGADNSTIELTONGFRNNATIDQWNAKNSDITVQYGG 120

QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 4

US-08-233-642A-57
Sequence 57, Application US/08233642A
GENERAL INFORMATION:
APPLICANT: Kay, William W.
APPLICANT: Collinson, S. Karen
APPLICANT: Clouthier, Sharon C.
APPLICANT: Doran, James L.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
BASED VACCINES
NUMBER OF SEQUENCES: 58
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed and Berry
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: U.S.A.
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233.642A
FILING DATE: 26-APR-1994
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: King, Joshua
REGISTRATION NUMBER: 35,570
REFERENCE/DOCKET NUMBER: 920043.403C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEEDANBERRY
INFORMATION FOR SEQ ID NO: 57:
SEQUENCE CHARACTERISTICS:
LENGTH: 151 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-233-642A-57

Query Match 87.9%; Score 687; DB 6; Length 151;
Best Local Similarity 90.1%; Pred. No. 3e-65;
Matches 136; Conservative 2; Mismatches 13; Indels 0; Gaps 0;


```
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-14

Query Match 78.3%; Score 612; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 3.6e-57;
Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALYDQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALQ 60

QY 61 LVTRVVTHEMAHAGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGOYGG 120
DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGOYDQ 120

QY 121 NNAALVNOTASDSSVMVROVGFNNATANQY 151
DB 121 LVTRVVTHEMAHASVMVROVGFNNATANQY 151

RESULT 9
US-09-543-407-24
; Sequence 24, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-24

Query Match 77.9%; Score 609; DB 19; Length 151;
Best Local Similarity 82.1%; Pred. No. 7.5e-57;
Matches 124; Conservative 4; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALYDQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALQ 60

QY 61 LVTRVVTHEMAHAGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGOYGG 120
DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVTRVVT 120
```

```
QY 121 NNAALVNOTASDSSVMVROVGFNNATANQY 151
DB 121 HEMAHANQASDSSVMVROVGFNNATANQY 151

RESULT 10
US-09-543-407-31
; Sequence 31, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Salmonella enteritidis
US-09-543-407-31

Query Match 77.4%; Score 605; DB 19; Length 131;
Best Local Similarity 89.3%; Pred. No. 1.7e-56;
Matches 117; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

QY 21 GVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALYDQLVTRVVTHEMAHAGYNGAD 80
DB 1 GVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALQSDARKSETTITQSGYNGAD 60

QY 81 VQGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGOYGGNNAALVNOTASDSSVMVROV 140
DB 61 VQGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGOYGGNNAALVNOTASDSSVMVROV 120

QY 141 GFGNNATANQY 151
DB 121 GFGNNATANQY 131

RESULT 11
US-09-543-407-22
; Sequence 22, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 22
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-22

Query Match 76.9%; Score 601; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 5.5e-56;
Matches 123; Conservative 4; Mismatches 24; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALYDQ 60
```

```
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDPSTLSIYQYGSANAALQ 60
QY 61 LVTRVVTHEMAHAGYNGADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITVGOYGG 120
Db 61 SDARKSETTTTQSGYNGADVGQADNYDQLVTRVVTHEMAHADQWNAKNSDITVGOYGG 120
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 12
US-09-543-407-30
; Sequence 30, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-30

Query Match 76.7%; Score 600; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 7e-56;
Matches 123; Conservative 4; Mismatches 24; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDPSTLSIYQYGSANAALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDPSTLSIYQYGSANAALQ 60
QY 61 LVTRVVTHEMAHAGYNGADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITVGOYGG 120
Db 61 SDARKSETTTTQSGYNGADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITVGOYGG 120
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 13
US-09-543-407-16
; Sequence 16, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-16

Query Match 77.7%; Score 600; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 7e-56;
Matches 123; Conservative 4; Mismatches 24; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDPSTLSIYQYGSANAALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDPSTLSIYQYGSANAALQ 60
QY 61 LVTRVVTHEMAHAGYNGADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITVGOYGG 120
Db 61 SDARKSETTTTQSGYNGADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITVGOYGG 120
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 14
US-09-543-407-7
; Sequence 7, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-7

Query Match 66.9%; Score 523; DB 19; Length 151;
Best Local Similarity 68.9%; Pred. No. 1.4e-47;
Matches 104; Conservative 18; Mismatches 29; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDPSTLSIYQYGSANAALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDPSTLSIYQYGSANAALQ 60
QY 61 LVTRVVTHEMAHAGYNGADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITVGOYGG 120
Db 61 TDARNSDLTITQGGGNGADVGQSDSSIDLTQRFNGSATLDQWNGKNSMTVVKQFGG 120
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
Db 121 GNGAAVDQTASNSVNVTVQVGFNNATANQY 151

RESULT 15
US-08-978-878-4
; Sequence 4, Application US/08978878
; GENERAL INFORMATION:
; APPLICANT: NORMARK, Staffan
; APPLICANT: OLSEN, Arne
; TITLE OF INVENTION: FIBRONECTIN BINDING PROTEIN AS WELL AS ITS PREPARATION
; FILE REFERENCE: 012889-081
; CURRENT FILING DATE: 1997-11-26
; CURRENT APPLICATION NUMBER: US/08/978,878
; EARLIER APPLICATION NUMBER: SE 8801723-1
; EARLIER FILING DATE: 1988-05-06
```

```
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-16

Query Match 73.8%; Score 577; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 2.1e-53;
Matches 122; Conservative 5; Mismatches 24; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDPSTLSIYQYGSANAALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDPSTLSIYQYGSANAALQ 60
QY 61 LVTRVVTHEMAHAGYNGADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITVGOYGG 120
Db 61 SDARKSETTTTQSGYNGADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITVGOYGG 120
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 14
US-09-543-407-7
; Sequence 7, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-7

Query Match 66.9%; Score 523; DB 19; Length 151;
Best Local Similarity 68.9%; Pred. No. 1.4e-47;
Matches 104; Conservative 18; Mismatches 29; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDPSTLSIYQYGSANAALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDPSTLSIYQYGSANAALQ 60
QY 61 LVTRVVTHEMAHAGYNGADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITVGOYGG 120
Db 61 TDARNSDLTITQGGGNGADVGQSDSSIDLTQRFNGSATLDQWNGKNSMTVVKQFGG 120
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
Db 121 GNGAAVDQTASNSVNVTVQVGFNNATANQY 151

RESULT 15
US-08-978-878-4
; Sequence 4, Application US/08978878
; GENERAL INFORMATION:
; APPLICANT: NORMARK, Staffan
; APPLICANT: OLSEN, Arne
; TITLE OF INVENTION: FIBRONECTIN BINDING PROTEIN AS WELL AS ITS PREPARATION
; FILE REFERENCE: 012889-081
; CURRENT FILING DATE: 1997-11-26
; CURRENT APPLICATION NUMBER: US/08/978,878
; EARLIER APPLICATION NUMBER: SE 8801723-1
; EARLIER FILING DATE: 1988-05-06
```

```
; EARLIER APPLICATION NUMBER: US 07/347,189
; EARLIER FILING DATE: 1989-05-04
; EARLIER APPLICATION NUMBER: US 07/789,437
; EARLIER FILING DATE: 1991-11-06
; EARLIER APPLICATION NUMBER: US 07/970,846
; EARLIER FILING DATE: 1992-11-03
; EARLIER APPLICATION NUMBER: US 08/187,865
; EARLIER FILING DATE: 1994-01-28
; EARLIER APPLICATION NUMBER: US 08/318,519
; EARLIER FILING DATE: 1994-10-05
; EARLIER APPLICATION NUMBER: US 08/495,959
; EARLIER FILING DATE: 1995-06-28
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-08-978-878-4

Query Match      66.5%; Score 520; DB 13; Length 151;
Best Local Similarity 68.2%; Pred. No. 2.9e-47;
Matches 103; Conservative 19; Mismatches 29; Indels 0; Gaps 0;

QY      1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGGNGSSGPDSTLSIYQYGSANAALYDQ 60
Db      1 MKLLKVAIAAIVFSGSAVAGVVPQYGGGNGGNGSSGPNSEINLYQYGGNSALALQ 60

QY      61 LVTRVVTHEMAHAGYCGADYCGGADNSTIELTQNGFNNATIDQWAKNSDITVQYGG 120
Db      61 TDARNSDLTITQHGNGGADYCGGSDSSIDLTORFGNSATLDQWNGKNSMTVKQFGG 120

QY      121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
Db      121 GNGAAVDQTASNSSVNVTVQVGFNNATANQY 151
```

Search completed: August 2, 2004, 15:26:46
Job time : 168.9 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:49:38 ; Search time 17.8 Seconds
(without alignments)
888.146 Million cell updates/sec

Title: US-09-543-407-26

Perfect score: 782

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVFQVGFNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 601315 seqs, 104695340 residues

Total number of hits satisfying chosen parameters: 601315

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending_parents_AA_New.*

1: /cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pep.*
2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep.*
3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep.*
4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep.*
5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep.*
6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep.*
7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	520	66.5	151	5	US-09-741-873C-4
2	442	56.5	131	5	US-09-741-873C-2
3	93	11.9	892	5	US-09-952-267B-5
4	93	11.9	892	6	US-10-872-768-5
5	93	11.9	892	6	US-10-872-769-5
6	92	11.8	295	6	US-10-425-115-312468
7	92	11.8	1459	6	US-10-482-706-129
8	91.5	11.7	891	7	US-60-566-425-574
9	91.5	11.7	891	7	US-60-576-812-606
10	91.5	11.7	894	7	US-60-566-425-570
11	91.5	11.7	894	7	US-60-566-425-571
12	91.5	11.7	894	7	US-60-566-425-575
13	91.5	11.7	894	7	US-60-566-425-578
14	91.5	11.7	894	7	US-60-576-812-602
15	91.5	11.7	894	7	US-60-576-812-603
16	91.5	11.7	894	7	US-60-576-812-607
17	91.5	11.7	894	7	US-60-576-812-610
18	91.5	11.7	898	6	US-10-170-205E-12329
19	91.5	11.7	898	7	US-60-566-425-577
20	91.5	11.7	898	7	US-60-576-812-609
21	89	11.4	197	6	US-10-425-115-304391
22	88.5	11.3	193	6	US-10-425-115-254240
23	88	11.3	1627	6	US-10-170-205E-16659
24	88	11.3	1905	1	PCT-US04-09388-9
25	87.5	11.2	234	6	US-10-767-701-45603
26	87	11.1	573	7	US-60-565-632-7907

ALIGNMENTS

RESULT 1

US-09-741-873C-4
; Sequence 4: Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873C-4

27	87	11.1	573	7	US-60-579-062-7907	Sequence 7907, Ap
28	85.5	10.9	258	6	US-10-425-115-300390	Sequence 300390,
29	85	10.9	147	6	US-10-425-115-193207	Sequence 193207,
30	85	10.9	412	7	US-60-565-632-7905	Sequence 7905, Ap
31	85	10.9	412	7	US-60-579-062-7905	Sequence 7905, Ap
32	85	10.9	841	7	US-60-565-632-7906	Sequence 7906, Ap
33	85	10.9	841	7	US-60-579-062-7906	Sequence 7906, Ap
34	83.5	10.7	443	6	US-10-100-683-7608	Sequence 7608, Ap
35	82	10.5	131	7	US-60-565-632-11109	Sequence 11109, A
36	82	10.5	131	7	US-60-579-062-11109	Sequence 11109, A
37	82	10.5	193	6	US-10-425-115-219256	Sequence 219256,
38	82	10.5	299	6	US-10-170-205E-35751	Sequence 35751, A
39	82	10.5	434	7	US-60-565-632-9168	Sequence 9168, Ap
40	82	10.5	434	7	US-60-579-062-9168	Sequence 9168, Ap
41	82	10.5	544	6	US-10-425-115-226699	Sequence 226699,
42	82	10.5	1303	6	US-10-170-205E-741	Sequence 741, Appl
43	82	10.5	1327	1	PCT-US04-02338-49	Sequence 49, Appl
44	82	10.5	1358	6	US-10-778-804-11	Sequence 11, Appl
45	81.5	10.4	179	6	US-10-767-701-35342	Sequence 35342, A

Query Match 66.5%; Score 520; DB 5; Length 151;
Best Local Similarity 68.2%; Pred. No. 5.5e-38;
Matches 103; Conservative 19; Mismatches 29; Indels 0; Gaps 0;

QY	1	MKLLKVAFAAIVVSGSALA	VVPGVGGGHHNGGSGSDSTLSIVYGSANAALYDQ	60
Db	1	MKLLKVAFAAIVVSGSALA	VVPGVGGGHHNGGSGNSGNSELNIYQYGGNSALALQ	60
QY	61	LVTRVVTTHMAGVCGADN	STIELTQNGFNNAITDQWNAKNSDITVGYGG	120
Db	61	TDARNSDLTITQHGCGG	ADVQGSDDSSIDLTRQFGNSATLDQWCKNSEMTVKQFGG	120
QY	121	NNAALVNQTASDSSVM	VFQVGFNNATANQY	151
Db	121	NGGAAVDQTASNSVNT	VQVGFNNATAHQY	151


```
; PRIOR FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 892
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-10-872-769-5

Query Match
Best Local Similarity 11.9%; Score 93; DB 6; Length 892;
Matches 41; Conservative 14; Mismatches 49; Indels 44; Gaps 9;

QY 28 GCGNHN-----GCGNSS---GPDSTLSIYQYGSANAALYDQLVTRVVTHEMAHAGY 75
DB 89 GGGDYNEAKGYSTVGGSSNTAKGKSTIGGDTNDANG-----TYSITGGY 137

QY 76 -----GNGADVGQADNSTI--ELTQNGFRNNAIDQWNAKNSDITVQYGV---GNNAL 125
DB 138 YSRAIGSSSTIGGGYNNQATGKSTVAGGRNN-----QATGNNSTVAGGSYNQATGNNSTV 193

QY 126 V-----NOTASDSSVMVRQVGFNNATAN 149
DB 194 AGGSHNQATGGSF---AAGVENKANAN 218

RESULT 6
US-10-425-115-312468
; Sequence 312468, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425.115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 312468
; LENGTH: 295
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_48027C.1.pep
US-10-425-115-312468

Query Match
Best Local Similarity 11.8%; Score 92; DB 6; Length 295;
Matches 37; Conservative 13; Mismatches 40; Indels 56; Gaps 5;

QY 26 WGGGNGHNGGNSGDPST-----LSIYQYGSANAALYDQLVTRVVTHEMAHAGYNGA 79
DB 138 YGGGGYSGGGYSSGGYAANGYGVGSGGNYNASGGYS-----GSDGYNGA 186

QY 80 DVGGQADNSTIELTQNGFRN-----NATIDQWNAK-----SDITVQYGV 119
DB 187 ASGGYANNLSSGYSNRYNTIGSSDNTGTYGYNPNYAGNYNTVGGSSSGTILGEFG 246

QY 120 GNNALYNQATSDSSVMVRQVGFNN 145
DB 247 GG-----GFGN 253

RESULT 7
US-10-482-706-129
; Sequence 129, Application US/10482706
; GENERAL INFORMATION:
; APPLICANT: James, Brian William
; APPLICANT: Marsh, Philip
; APPLICANT: Hampshire, Tobias
; TITLE OF INVENTION: Mycobacterial Antigens Expressed During Latency

; FILE REFERENCE: 1581.1030000
; CURRENT APPLICATION NUMBER: US/10/482.706
; CURRENT FILING DATE: 2004-01-02
; PRIOR APPLICATION NUMBER: PCT/GB02/03052
; PRIOR FILING DATE: 2002-07-04
; PRIOR APPLICATION NUMBER: GB 0116385.6
; PRIOR FILING DATE: 2001-07-04
; PRIOR APPLICATION NUMBER: GB 0123993.8
; PRIOR FILING DATE: 2001-10-05
; NUMBER OF SEQ ID NOS: 292
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 129
; LENGTH: 1459
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-10-482-706-129

Query Match
Best Local Similarity 11.8%; Score 92; DB 6; Length 1459;
Matches 38; Conservative 20; Mismatches 76; Indels 18; Gaps 5;

QY 7 AAFAAIVVSGSALAGVVPQGGGHNHGGNSGDPSTLSIYQYGSANAALYDQLVTRV 56
DB 722 SGFGNVGSGSGFWNIAGNLGNSGFLNVGLTSG-----ILNFGTVSGLYNTSTLGLA 775

QY 67 THEMAHAGYNG-----ADVGQADNSTIELTQNGFRNNAIDQWNAKNSDI-----TVGQ 117
DB 776 TSAF-HSGVGNLDSQLAGPMNAAGT--LPNFGANDGTNLGNANLGDYNVGSGVGS 832

QY 118 YGNNALYNQATSDSSVMVRQVGFNNATAN 149
DB 833 YNFGSGNIGNSGFGNIGNSNPNFGNFGVGSNN 864

RESULT 8
US-60-566-425-574
; Sequence 574, Application US/60566425
; GENERAL INFORMATION:
; APPLICANT: DOMON, Bruno et al.
; TITLE OF INVENTION: COLON DISEASE TARGETS AND USES THEREOF
; FILE REFERENCE: CL001522-PROV
; CURRENT APPLICATION NUMBER: US/60/566.425
; CURRENT FILING DATE: 2004-04-30
; NUMBER OF SEQ ID NOS: 1422
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 574
; LENGTH: 891
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-566-425-574

Query Match
Best Local Similarity 11.7%; Score 91.5; DB 7; Length 891;
Matches 35; Conservative 15; Mismatches 32; Indels 47; Gaps 7;

QY 27 GGGG-----NHNGGNSGDPSTLSIYQYGSANAALYDQLVTRVVTHEMAHAGYCN 77
DB 792 GGGGSDYNEKFNYSGGSGSGNS-----YSGGASY-----NPGSHGGYGG 835

QY 78 GADVGGQADNSTIELTQNGFRNNAIDQWNAKNSDITVQYCGNNAALYNQATSDSSVMV 137
DB 836 GSGGG-----SSYQKQGGYSQS-----NYSNPGS-----GQ-----NYSGPSSYSQS 873

QY 138 RQVGFNNNA 146
DB 874 SQGGYGRNA 882

RESULT 9
US-60-576-812-606
; Sequence 606, Application US/60576812
; GENERAL INFORMATION:
; APPLICANT: DOMON, Bruno
```

us-09-543-407-26.rapn

Tue Aug 3 10:54:43 2004

```
; TITLE OF INVENTION: COLON DISEASE TARGETS AND USES THEREOF
; FILE REFERENCE: CL001527PROV
; CURRENT APPLICATION NUMBER: US/60/576,812
; CURRENT FILING DATE: 2004-06-04
; NUMBER OF SEQ ID NOS: 1501
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 606
; LENGTH: 891
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-576-812-606

Query Match      11.7%; Score 91.5; DB 7; Length 891;
Best Local Similarity 27.1%; Pred. No. 7.7;
Matches 35; Conservative 15; Mismatches 32; Indels 47; Gaps 7;

QY 27 GGGG-----NHHGGGSSGPDSTLSIYQYGSANAALYDQLVTRVVTTHMAHAGYGN 77
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 792 GGGGSDYNYESKFNSGGRSGGNS-----YSGGASY-----NPGSHGGYGG 835

QY 78 GADVGGGADNSTIETQNGFRNNATIDOWNAKNSDITVQYGGNNAALVQNTASDSSVMV 137
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 836 GSGGG-----SSYQKGQGGYSQS-----NYSNPGS-----GQ-----NYSGPPSSYSQS 873

QY 138 RQVGFGNNA 146
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 874 SQGGYGRNA 882

RESULT 10
US-60-566-425-570
; Sequence 570, Application US/60566425
; GENERAL INFORMATION:
; APPLICANT: DOMON, Bruno et al.
; TITLE OF INVENTION: COLON DISEASE TARGETS AND USES THEREOF
; FILE REFERENCE: CL001522-PROV
; CURRENT APPLICATION NUMBER: US/60/566,425
; CURRENT FILING DATE: 2004-04-30
; NUMBER OF SEQ ID NOS: 1422
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 570
; LENGTH: 894
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-566-425-570

Query Match      11.7%; Score 91.5; DB 7; Length 894;
Best Local Similarity 27.1%; Pred. No. 7.7;
Matches 35; Conservative 15; Mismatches 32; Indels 47; Gaps 7;

QY 27 GGGG-----NHHGGGSSGPDSTLSIYQYGSANAALYDQLVTRVVTTHMAHAGYGN 77
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 795 GGGGSDYNYESKFNSGGRSGGNS-----YSGGASY-----NPGSHGGYGG 838

QY 78 GADVGGGADNSTIETQNGFRNNATIDOWNAKNSDITVQYGGNNAALVQNTASDSSVMV 137
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 839 GSGGG-----SSYQKGQGGYSQS-----NYSNPGS-----GQ-----NYSGPPSSYSQS 876

QY 138 RQVGFGNNA 146
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 877 SQGGYGRNA 885

RESULT 11
US-60-566-425-571
; Sequence 571, Application US/60566425
; GENERAL INFORMATION:
; APPLICANT: DOMON, Bruno et al.
; TITLE OF INVENTION: COLON DISEASE TARGETS AND USES THEREOF
; FILE REFERENCE: CL001522-PROV
; CURRENT APPLICATION NUMBER: US/60/566,425
; CURRENT FILING DATE: 2004-04-30
; NUMBER OF SEQ ID NOS: 1422
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 578
; LENGTH: 894
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-566-425-571

Query Match      11.7%; Score 91.5; DB 7; Length 894;
Best Local Similarity 27.1%; Pred. No. 7.7;
Matches 35; Conservative 15; Mismatches 32; Indels 47; Gaps 7;

QY 27 GGGG-----NHHGGGSSGPDSTLSIYQYGSANAALYDQLVTRVVTTHMAHAGYGN 77
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 795 GGGGSDYNYESKFNSGGRSGGNS-----YSGGASY-----NPGSHGGYGG 838

QY 78 GADVGGGADNSTIETQNGFRNNATIDOWNAKNSDITVQYGGNNAALVQNTASDSSVMV 137
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 839 GSGGG-----SSYQKGQGGYSQS-----NYSNPGS-----GQ-----NYSGPPSSYSQS 876

QY 138 RQVGFGNNA 146
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 877 SQGGYGRNA 885

RESULT 12
US-60-566-425-575
; Sequence 575, Application US/60566425
; GENERAL INFORMATION:
; APPLICANT: DOMON, Bruno et al.
; TITLE OF INVENTION: COLON DISEASE TARGETS AND USES THEREOF
; FILE REFERENCE: CL001522-PROV
; CURRENT APPLICATION NUMBER: US/60/566,425
; CURRENT FILING DATE: 2004-04-30
; NUMBER OF SEQ ID NOS: 1422
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 575
; LENGTH: 894
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-566-425-575

Query Match      11.7%; Score 91.5; DB 7; Length 894;
Best Local Similarity 27.1%; Pred. No. 7.7;
Matches 35; Conservative 15; Mismatches 32; Indels 47; Gaps 7;

QY 27 GGGG-----NHHGGGSSGPDSTLSIYQYGSANAALYDQLVTRVVTTHMAHAGYGN 77
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 795 GGGGSDYNYESKFNSGGRSGGNS-----YSGGASY-----NPGSHGGYGG 838

QY 78 GADVGGGADNSTIETQNGFRNNATIDOWNAKNSDITVQYGGNNAALVQNTASDSSVMV 137
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 839 GSGGG-----SSYQKGQGGYSQS-----NYSNPGS-----GQ-----NYSGPPSSYSQS 876

QY 138 RQVGFGNNA 146
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 877 SQGGYGRNA 885

RESULT 13
US-60-566-425-578
; Sequence 578, Application US/60566425
; GENERAL INFORMATION:
; APPLICANT: DOMON, Bruno et al.
; TITLE OF INVENTION: COLON DISEASE TARGETS AND USES THEREOF
; FILE REFERENCE: CL001522-PROV
; CURRENT APPLICATION NUMBER: US/60/566,425
; CURRENT FILING DATE: 2004-04-30
; NUMBER OF SEQ ID NOS: 1422
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 578
; LENGTH: 894
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-566-425-578
```


This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:39:53 ; Search time 9.4 Seconds
(without alignments)
1545.204 Million cell updates/sec

Title: US-09-543-407-26

Perfect score: 782
Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVFGNNATANQY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78.*

- 1: PIR1.*
- 2: PIR2.*
- 3: PIR3.*
- 4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	692	88.5	151	2 JC6039	fimbrin protein ag
2	692	88.5	151	2 A70635	major curlin chain
3	523	66.9	151	2 S70788	curlin protein csg
4	501.5	64.1	152	2 D90806	curlin major subun
5	501.5	64.1	152	2 H85665	hypothetical prote
6	114.5	14.6	145	2 AD3143	conserved hypotet
7	114.5	14.6	145	2 H88144	hypothetical prote
8	113	14.5	2174	2 E95965	hypothetical glyci
9	112	14.3	151	2 S70787	curlin nucleator p
10	112	14.3	151	2 C90806	minor curlin subun
11	112	14.3	151	2 G85665	curlin minor chain
12	107.5	13.7	1322	2 S07053	ice nucleation pro
13	107	13.7	151	2 JC6040	fimbrin protein ag
14	107	13.7	151	2 AH0635	nucleation compone
15	106	13.6	1034	2 JC2143	ice nucleation act
16	105.5	13.5	645	2 F70625	probable PPE prote
17	105	13.4	590	2 E70946	probable PPE prote
18	104.5	13.4	1258	2 JQ0188	ice nucleation pro
19	102.5	13.1	552	2 D70604	probable PPE prote
20	102.5	13.1	1567	2 S11672	ice nucleation pro
21	102.5	13.1	1655	2 E97935	hypothetical prote
22	100	12.8	3300	2 D70575	probable PPE prote
23	97.5	12.5	1028	2 A56038	DNA-binding protei
24	97.5	12.5	1213	2 S16356	ovo protein - frui
25	97	12.4	615	2 E70663	probable PPE prote
26	96.5	12.3	447	2 G84687	probable disease r
27	95.5	12.2	1651	2 JC1340	outer membrane pro
28	95	12.1	652	2 E7857	cell surface antiq
29	94	12.0	599	2 B42049	leishmanolysin (EC

30	94	12.0	599	2 A44951	leishmanolysin (EC
31	93	11.9	434	2 E70768	hypothetical glyci
32	92.5	11.8	575	2 S35327	protein kinase sgg
33	92.5	11.8	639	2 C42049	leishmanolysin (EC
34	92.5	11.8	963	2 B70524	probable PPE prote
35	92.5	11.8	1053	2 B70987	probable PPE prote
36	92	11.8	354	2 B70663	probable PPE prote
37	92	11.8	1436	2 B70520	probable PPE prote
38	92	11.8	1748	2 S42136	cnjB protein - Tet
39	91.5	11.7	582	2 F70675	probable PPE prote
40	91	11.6	678	2 A70762	probable PPE prote
41	90.5	11.6	1200	1 SNFSO	ice nucleation pro
42	90.5	11.6	3716	2 E70969	probable PPE prote
43	90	11.5	586	2 T26667	hypothetical prote
44	90	11.5	1390	2 T14004	trfA protein - sli
45	89.5	11.4	1317	2 A54831	nuclear pore compl

ALIGNMENTS

RESULT 1

JC6039
fimbrin protein agfA precursor - Salmonella enteritidis
C:Species: Salmonella enteritidis
C>Date: 31-Dec-1996 #sequence revision 31-Dec-1996 #text_change 08-Oct-1999
C:Accession: JC6039; PC6015; A44898
R:Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Baner, P.A.; Kay, W.W.
J. Bacteriol. 178, 662-667, 1996
A>Title: Salmonella enteritidis agfBAC operon encoding thin, aggregative fimbrinae.
A:Reference number: JC6039; MUID:96146512; PMID:8550497
A:Accession: JC6039
A:Molecule type: DNA
A:Residues: 1-151 <COL>
A:Cross-references: GB:U43280; NID:g1184712; PIDN:AAC43599.1; PID:g1184714
A:Accession: PC6015
A:Molecule type: protein
A:Residues: 21-52 <CO2>
A:Experimental source: strain 27655-3b
A>Note: the authors translated the codon ACG for residue 44 as Ile
R:Collinson, S.K.; Emody, L.; Muller, K.H.; Trust, T.J.; Kay, W.W.
J. Bacteriol. 173, 4773-4781, 1991
A>Title: Purification and characterization of thin, aggregative fimbrinae from Salmonella
A:Reference number: A44898; MUID:91310586; PMID:1677357
A:Contents: 27655
A:Accession: A44898
A>Status: preliminary
A:Molecule type: protein
A:Residues: 21-33 <CO3>
A:Notes: sequence extracted from NCBI backbone (NCBIP:45936)
C:Genetics:
A:Gene: agfA
C:Function:
A>Description: major component of thin aggregative fimbrinae
A>Note: fimbrinae bind to fibronectin, plasminogen, tissue plasminogen activator
C:Keywords: fimbrina
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-151/Product: fimbrin protein agfA #status experimental <MAT>

Query Match 88.5%; Score 692; DB 2; Length 151;
Best Local Similarity 90.7%; Pred. No. 3.6e-51;
Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

QY	1	MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSSGGPDSLTLSIYQGSANNAALYDQ	60
DB	1	MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSSGGPDSLTLSIYQGSANNAALQ	60
QY	61	LVTRVVVTHEMAHAGYNGADVGQGDADNSTIELTQNGFNPNATIDOWNAKNSDITVGOYGG	120
DB	61	SPARKSETTICSGYNGADVGQGDADNSTIELTQNGFNPNATIDOWNAKNSDITVGOYGG	120
QY	121	NNAALVNQATSDSSVWVRQVFGNNATANQY	151

Db 121 NNAALVNOTASDSSVMVRQVFGNNATANQY 151

RESULT 2

At0635

major curlin chain precursor [imported] - Salmonella enterica subsp. enterica serovar Typhimurium

C:Species: Salmonella enterica subsp. enterica serovar Typhi

A:Note: this species has also been called Salmonella typhi

C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002

C:Accession: At0635

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.

A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhimurium

A:Reference number: AB0502; PMID:21534947; PMID:11677608

A:Accession: At0635

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-151 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD08268.1; PID:g16502315; GSPDB:GN00176

C:Genetics:

A:Gene: STY1181

Query Match 88.5%; Score 692; DB 2; Length 151;

Best Local Similarity 90.7%; Pred. No. 3.6e-51;

Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVSGSALAGVVPQGGGNGHNGSGSPDSTLSIYQYGSANAALYDQ 60

Db 1 MKLLKVAFAAIVSGSALAGVVPQGGGNGHNGSGSPDSTLSIYQYGSANAALQ 60

QY 61 LVTRVVTHEMAHAGVNGADVCGQADNSTIELTQNGFNPNATIDQWNAKNSDITVQYGG 120

Db 61 SPARKSETTITQSGYNGADVCGQADNSTIELTQNGFNPNATIDQWNAKNSDITVQYGG 120

QY 121 NNAALVNOTASDSSVMVRQVFGNNATANQY 151

Db 121 NNAALVNOTASDSSVMVRQVFGNNATANQY 151

RESULT 3

S70788

curlin protein csGA precursor - Escherichia coli (strain K-12)

N:Alternate names: csGA protein; major curlin protein

C:Species: Escherichia coli

C:Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 01-Mar-2002

C:Accession: S70788; G64846; S31202; S34559

R:Hammar, M.; Arngqvist, A.; Bian, Z.; Olsen, A.; Normark, S.

Mol. Microbiol. 18, 661-670, 1995

A:Title: Expression of two csG operons is required for production of fibronectin- and C

A:Reference number: S70783; PMID:96414468; PMID:8817489

A:Accession: S70788

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-151 <HAM>

A:Cross-references: EMBL:X90754; NID:g1147558; PIDN:CAA62282.1; PID:g1147564

A:Experimental source: strain K12, substrain W3110

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co

A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; PMID:97426617; PMID:9278503

A:Accession: G64846

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-151 <BLAT>

A:Cross-references: GB:AB0000205; GB:U00096; NID:g1787265; PIDN:AAC74126.1; PID:g1787279;

A:Experimental source: strain K-12, substrain MG1655

R:Olsen, A.; Arngqvist, A.; Hammar, M.; Sukupolvi, S.; Normark, S.

Mol. Microbiol. 7, 523-536, 1993

A:Title: The RpoS sigma factor relieves H-NS-mediated transcriptional repression of csGA

A:Reference number: S31202; PMID:93211294; PMID:8459772

A:Accession: S31202

A:Molecule type: DNA

A:Residues: 1-6 'V', 8-151 <OLS1>

A:Cross-references: EMBL:L04979

A:Accession: S34560

A:Molecule type: protein

A:Residues: 21-42; 44-50 <OLS2>

R:Olsen, A.N.; Arngqvist, A.M.

submitted to the EMBL Data Library, October 1992

A:Reference number: S34559

A:Accession: S34559

A:Molecule type: DNA

A:Residues: 1-133, 'RQDSGWLW' <OLS3>

A:Cross-references: EMBL:L04979; NID:g290424; PIDN:AAA23616.1; PID:g290425

A:Experimental source: strain K-12, substrain W3110

C:Genetics:

A:Gene: csGA

A:Map position: 23.15

C:Function:

A:Description: major component of wild-type curli; interaction between csGA and csGB tr

A:Note: curli are thin, coiled fibers expressed on the surface of Escherichia coli that

and H-kininogen; in the absence of csGA, csGB can self-assemble into polymers

F:1-20/Domain: signal sequence #status predicted <SIG>

F:21-151/Product: curlin #status experimental <MAT>

Query Match 66.9%; Score 523; DB 2; Length 151;

Best Local Similarity 68.9%; Pred. No. 5.3e-37;

Matches 104; Conservative 18; Mismatches 29; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVSGSALAGVVPQGGGNGHNGSGSPDSTLSIYQYGSANAALYDQ 60

Db 1 MKLLKVAFAAIVSGSALAGVVPQGGGNGHNGSGSPDSTLSIYQYGSANAALQ 60

QY 61 LVTRVVTHEMAHAGVNGADVCGQADNSTIELTQNGFNPNATIDQWNAKNSDITVQYGG 120

Db 61 TDARNSDLTITQHGNGGADVCGQSDSSIDITQRFNGNSATLDQWNGKNSMTVQYFGG 120

QY 121 NNAALVNOTASDSSVMVRQVFGNNATANQY 151

Db 121 GNGAAVDQTASNSVNVTVQVFGNNATAHQY 151

RESULT 4

D90806

curlin major subunit csGA [imported] - Escherichia coli (strain O157:H7, substrain RIMD

C:Species: Escherichia coli

C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001

C:Accession: D90806

R:Hayaishi, T.; Makino, K.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.

gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingagawa, H.

DNA Res. 9, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen

A:Reference number: A99629; PMID:11258796

A:Accession: D90806

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-152 <HAY>

A:Cross-references: GB:BA000007; PIDN:BA834843.1; PID:g13360880; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RIMD 050952

C:Genetics:

A:Gene: Ecs1420

Query Match 64.1%; Score 501.5; DB 2; Length 152;

Best Local Similarity 67.1%; Pred. No. 3.4e-35;

Matches 102; Conservative 19; Mismatches 30; Indels 1; Gaps 1;

QY 1 MKLLKVAFAAIVSGSALAGVVPQW-GGCGNHNGSGSGPDSTLSIYQYGSANAALYD 59

Db 1 MKLLKVAFAAIVSGSALAGVVPQYGGGNGHNGSGNSPNSLNIYQYGGNSALAL 60

QY 60 QLTVRVVTHEMAHAGVNGADVCGQADNSTIELTQNGFNPNATIDQWNAKNSDITVQYGG 119

Db 61 QADAENSDLTITQHGCGNGADVGQSGDDSSIDLTRQFGNSATLQWNGKDSHTVKQFG 120
 QY 120 GNNALVNQTASDSSVMVRQVGFGNNTATNQY 151
 Db 121 GGNGAAVDQTASNTVNVTQVGFGNNTATNQY 152

RESULT 5

H85665
 hypothetical protein csgA [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
 C:Species: Escherichia coli
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 C:Accession: H85665
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
 Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
 Nature 409, 529-533, 2001
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A:Reference number: A95480; MUID:21074935; PMID:11206551
 A:Accession: H85665
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-152 <STO>
 A:Cross-references: GB:AB005174; NID:g12514574; PIDN:AAG55788.1; GSPDB:GN00145; UWGP:Z16
 A:Experimental source: strain O157:H7, substrain EDL933
 C:Genetics:
 A:Gene: csgA

Query Match 64.1%; Score 501.5; DB 2; Length 152;
 Best Local Similarity 67.1%; Pred. No. 3.4e-35;
 Matches 102; Conservative 19; Mismatches 30; Indels 1; Gaps 1;

QY 1 MLLKVAAPAIIVVSGSALAGVVPQW-GGGNHNGGNSGPDSTLSIYQGSANAALYD 59
 Db 1 MLLKVAAPAIIVVSGSALAGVVPQWGGGGNGGNSGPDSTLSIYQGSANAL 60
 QY 60 QLVTRVVTHEMAHAGYNGADVGOGADNSTIELTQNGFRNNATIDQWNAKNSDITVGOYG 119
 Db 61 QADARNSDLTITQHGCGNGADVGQSGDDSSIDLTRQFGNSATLQWNGKDSHTVKQFG 120
 QY 120 GNNALVNQTASDSSVMVRQVGFGNNTATNQY 151
 Db 121 GGNGAAVDQTASNTVNVTQVGFGNNTATNQY 152

RESULT 6

AD3143
 conserved hypothetical protein Atu4768 [imported] - Agrobacterium tumefaciens (strain C58)
 C:Species: Agrobacterium tumefaciens
 C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
 C:Accession: AD3143
 R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, H.
 erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
 i, Karp, P.; Romero, P.; Zhang, S.
 Science 294, 2317-2323, 2001
 A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
 ster, E.W.

A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
 A:Reference number: AB2577; MUID:21608550; PMID:11743193
 A:Accession: AD3143
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-145 <KUR>
 A:Cross-references: GB:AE008689; PIDN:AAL45562.1; PID:g17743277; GSPDB:GN00187
 A:Experimental source: strain C58 (Dupont)
 C:Genetics:

A:Gene: Atu4768
 A:Map position: linear chromosome

Query Match 14.6%; Score 114.5; DB 2; Length 145;
 Best Local Similarity 25.2%; Pred. No. 0.0088;
 Matches 39; Conservative 26; Mismatches 59; Indels 31; Gaps 4;

QY 3 LLKVAAPAIIVVSGSALAGVVPQWGG-----GGNHNGGNSGPDSTLSIYQGSANA 55
 Db 1 MIRKSFIALVALVGLSAAAPAMANDVRIEQVGSNSAGGAQEGYGNRIITYQNGYN- 59
 QY 56 ALYDQLVTRVVTHEMAHAGYNGADVGOGADNSTIELTQNGFRNNATIDQWNAKNSDITV 115
 Db 60 -----RIVGHQY---GRHNLSAVGQEGHDNYGSTTQNGNRNVAGI----- 96
 QY 116 GOYGGNNAALVNQTASDSSVMVRQVGFGNNTATNQ 150
 Db 97 GQFGSNHTTILTDGNGNIAAGVQVGRGCSANVSQ 131

RESULT 7

H98144
 hypothetical protein AGE_L_228 [imported] - Agrobacterium tumefaciens (strain C58, Cere
 C:Species: Agrobacterium tumefaciens
 C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
 C:Accession: H98144
 R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman
 A.; Liu, F.; Wollam, C.; Allinger, M.; Doughy, D.; Scott, C.; Lappas, C.; Markelz, B
 Science 294, 2323-2328, 2001
 A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tu
 A:Reference number: A97359; MUID:21608551; PMID:11743194
 A:Accession: H98144
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-145 <KUR>
 A:Cross-references: GB:AE007870; PIDN:AAK8682.1; PID:g15158413; GSPDB:GN00170
 C:Genetics:
 A:Gene: AGE_L_228
 A:Map position: linear chromosome

Query Match 14.6%; Score 114.5; DB 2; Length 145;
 Best Local Similarity 25.2%; Pred. No. 0.0088;
 Matches 39; Conservative 26; Mismatches 59; Indels 31; Gaps 4;

QY 3 LLKVAAPAIIVVSGSALAGVVPQWGG-----GGNHNGGNSGPDSTLSIYQGSANA 55
 Db 1 MIRKSFIALVALVGLSAAAPAMANDVRIEQVGSNSAGGAQEGYGNRIITYQNGYN- 59
 QY 56 ALYDQLVTRVVTHEMAHAGYNGADVGOGADNSTIELTQNGFRNNATIDQWNAKNSDITV 115
 Db 60 -----RIVGHQY---GRHNLSAVGQEGHDNYGSTTQNGNRNVAGI----- 96
 QY 116 GOYGGNNAALVNQTASDSSVMVRQVGFGNNTATNQ 150
 Db 97 GQFGSNHTTILTDGNGNIAAGVQVGRGCSANVSQ 131

RESULT 8

E95965
 hypothetical glycine-rich protein [imported] - Sinorhizobium meliloti (strain 1021) mag
 C:Species: Sinorhizobium meliloti
 C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
 C:Accession: E95965
 R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Herna
 Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
 A:Title: The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-fixing end
 A:Reference number: A95842; MUID:21396508; PMID:11481431
 A:Accession: E95965
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-2174 <KUR>

A:Cross-references: PIDN:CAC49389.1; PID:g15140875; GSPDB:GN00167
 A:Experimental source: strain 1021, megaplasmid pSymb
 R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler
 pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.
 L.; Hyman, R.W.; Jones, T.
 Science 293, 668-672, 2001
 A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Ielaure
 Hebault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K
 A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.

A:Reference number: A96039; MUID:21368234; PMID:11474104

A:Contents: annotation

C:Genetics:

A:Gene: SMB21548

A:Genome: plasmid

Query Match 14.5%; Score 113; DB 2; Length 2174;
Best Local Similarity 27.0%; Pred. No. 0.23;
Matches 40; Conservative 21; Mismatches 51; Indels 36; Gaps 7;
QY 11 AIVVSGSALAGVWPQ--WGGGNGNNGGNSGPPSTLSIVQYGSANAALYDQLVTRVWTH 68
DB 693 AATAGAGAGTILAQSIGGGGNG--CGNATGGDAGFGSFQIGGGGG----- 737
QY 69 EMAHAGYNGGADVG-----QGADNSTI--ELTQNGFRNATIDOWNAK--NSDITV 115
DB 738 ---GGYANTANVGFKGLTLTQGHAAAGIVAQSVGGGGTGTASSYAGIGFTASVAV 793
QY 116 GQYGGNAA--LVNQTSADSSVMVRQVG 141
DB 794 GGTGGNGGAGGEVSVSLTDSAIRTGGGG 821

RESULT 9

S70787
curlin nucleator protein csb precursor - Escherichia coli (strain K-12)
N:Alternate names: csb protein; curlin nucleation component; minor curlin protein
C:Species: Escherichia coli
C>Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 01-Mar-2002
C:Accession: S70787; 264846
R:Hammar, M.; Arngvist, A.; Bian, Z.; Olsen, A.; Normark, S.
Mol. Microbiol. 18, 661-670, 1995
A:Title: Expression of two csb operons is required for production of fibronectin- and Cc
A:Reference number: S70783; MUID:96414468; PMID:8817489
A:Accession: S70787
A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-151 <HAM>

A:Cross-references: EMBL:X90754; NID:g1147558; PIDN:CAA62281.1; PID:g1147563

A:Experimental source: strain K12, substrain W3110

A>Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co

A.: Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1433-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: F64846

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-151 <BLAT>

A:Cross-references: GB:AE000205; GB:U00096; NID:g1787265; PIDN:AAC74125.1; PID:g1787278

A:Experimental source: strain K-12, substrain MG1655

C:Genetics:

A:Gene: csbB

A:Map position: 23.15

C:Function:

A:Description: minor component of wild-type curli; interaction between CsgA and CsgB tri

A>Note: curli are thin, coiled fibers expressed on the surface of Escherichia coli that

and H-kininogen; in the absence of CsgA, CsgB can self-assemble into polymers

F1-21/Domain: signal sequence #status predicted <Sig>

F122-151/Product: minor curlin chain #status predicted <Mat>

Query Match 14.3%; Score 112; DB 2; Length 151;
Best Local Similarity 29.8%; Pred. No. 0.015;
Matches 34; Conservative 17; Mismatches 49; Indels 14; Gaps 3;

QY 51 GSANAALYDQLVLT--RVVTHEMAHAGYNGGADVGGQADNSTIELTQNGFR----- 98

DB 17 GIAAAGYDLANSEYNFVAVNELSKSFNQAAIIGQAGTNNSAQLRQGGSKLLAVVAQEGS 76

QY 99 -NNATIDOWNAKNSDITVCGYGGNNAALVNTQASDSSVMVRQVGFNNATANQY 151

DB 77 SNEAKIDQTDYNL-AYIDQASANDASISQAGYNTAMIIQKSGNKANITQY 129

RESULT 10

C90806
minor curlin subunit precursor CsgB [imported] - Escherichia coli (strain O157:H7, subst
C:Species: Escherichia coli
C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: C90806

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.

Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene

A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: C90806

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-151 <HAY>

A:Cross-references: GB:BA000007; PIDN:BA34842.1; PID:g13360879; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RIMD 050952

C:Genetics:

A:Gene: ECs1419

Query Match 14.3%; Score 112; DB 2; Length 151;
Best Local Similarity 29.8%; Pred. No. 0.015;
Matches 34; Conservative 17; Mismatches 49; Indels 14; Gaps 3;

QY 51 GSANAALYDQLVLT--RVVTHEMAHAGYNGGADVGGQADNSTIELTQNGFR----- 98

DB 17 GIAAAGYDLANSEYNFVAVNELSKSFNQAAIIGQAGTNNSAQLRQGGSKLLAVVAQEGS 76

QY 99 -NNATIDOWNAKNSDITVCGYGGNNAALVNTQASDSSVMVRQVGFNNATANQY 151

DB 77 SNEAKIDQTDYNL-AYIDQASANDASISQAGYNTAMIIQKSGNKANITQY 129

RESULT 11

G85665
curlin minor chain precursor, CsgA homolog [imported] - Escherichia coli (strain O157:H7

C:Species: Escherichia coli

C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001

C:Accession: G85665

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew

iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca

Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: G85665

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-151 <STO>

A:Cross-references: GB:AE005174; NID:g12514573; PIDN:AAG55787.1; GSPDB:GN00145; UWGP:Z16

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: csbB

Query Match 14.3%; Score 112; DB 2; Length 151;
Best Local Similarity 29.8%; Pred. No. 0.015;
Matches 34; Conservative 17; Mismatches 49; Indels 14; Gaps 3;

QY 51 GSANAALYDQLVLT--RVVTHEMAHAGYNGGADVGGQADNSTIELTQNGFR----- 98

DB 17 GIAAAGYDLANSEYNFVAVNELSKSFNQAAIIGQAGTNNSAQLRQGGSKLLAVVAQEGS 76

QY 99 -NNATIDOWNAKNSDITVCGYGGNNAALVNTQASDSSVMVRQVGFNNATANQY 151

DB 77 SNEAKIDQTDYNL-AYIDQASANDASISQAGYNTAMIIQKSGNKANITQY 129

RESULT 12

S07053
ice nucleation protein inaA - Erwinia ananas

C:Species: Erwinia ananas

C>Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 26-Aug-1999

C;Accession: S07053
R;Abe, K.; Watabe, S.; Emori, Y.; Watanabe, M.; Arai, S.
FEBS Lett. 258, 297-300; 1989
A;Title: An ice nucleation active gene of *Ewinia ananas*. Sequence similarity to those
A;Reference number: S07053; MUID:90052494; PMID:2595095
A;Accession: S07053
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-1322 <ABE>
A;Cross-references: GB:X17316; NID:G9296095; PIDN:CAA35194.1; PID:G9296096
C;Superfamily: ice nucleation protein

```

RESULT 13
JC6040
fimbriin protein agfB precursor - Salmonella enteritidis
C:Species: Salmonella enteritidis
C:Date: 31-Dec-1996 #sequence_revision 31-Dec-1996 #text_change 08-Oct-1999
C:Accession: JC6040
R:Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Banser, P.A.; Kay, W.W.
J. Bacteriol. 178, 662-667, 1996
A:Title: Salmonella enteritidis agfBAC operon encoding thin, aggregative fimbriae.
A:Reference number: JC6039; MUID:96148512; PMID:8550497
A:Accession: JC6040
A:Molecule type: DNA
A:Residues: 1-151 <COL>
A:Cross-references: GB:U43280; NID:g1184712; PIDN:AAC43598.1; PID:g1184713
A:Experimental source: strain 276755-3b
C:Genetics:
A:Gene: agfB
C:Function:
A:Description: minor component of thin aggregative fimbriae
A:Note: fimbriae bind to fibronectin, plasminogen, tissue plasminogen activator
C:Keywords: fimbria
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-151/Product: fimbriin protein agfB #status predicted <MAT>

```

RESULT 14
AH0635
nucleation component of curlin monomers [imported] - Salmonella enterica subsp. enterica
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: This species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AH0635

R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, E.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moulie, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.
A;Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* sero
A:Reference number: AB0502; MUID:21534947; PMID:116577608
A;Accession:AH0635
A;Status: Preliminary
A;Molecule type: DNA
A;Residues: 1-151 <PAR>
A;Cross-references: GB:AL513362; FIDN:CAP09267.1; PID:gl502314; GSPDB:GN00176
C;Genetics: +
A;Gene: STY1180

```

Query Match      13.7%; Score 107; DB 2; Length 151;
Best Local Similarity 30.4%; Pred. No. 0.039;
Matches 35; Conservative 17; Mismatches 47; Indels 16; Gaps 5;

QY      51  GSANAALYD--QLVTRVVTHMAHAGYGCADYVQ--CADNST-----HETLQNGF 97
Db      17  GIATATNYDLAREYNFAVNNELSKSEFNQAATIGCVGTDSARVQRQEGSKLLSVISQEGE 76

QY      98  RNNATIDQWNAKSDIT-VQYQYGNNAALVNTASDSSVMVRQVGFQGNNTATNOY 151
Db      77  NNRKAVDQ--AGNYNFAYIETQVANDASISQSAYGNSAAIIQKSGNKANITQY 129

RESULT 15
JC2143
ice nucleation active protein - Erwinia uredovora (strain KUIN-3)
C:Species: Erwinia uredovora
C/Date: 30-Sep-1993 #sequence_revision 20-Aug-1994 #text_change 17-Mar-1999
C/Accession: JC2143
R/Michigami, Y.; Watabe, S.; Abe, K.; Obata, H.; Arai, S.
BioSci. Biotechnol. Biochem. 58, 762-764, 1994
A/Title: Cloning and sequencing of an ice nucleation active gene of Erwinia uredovora.
A/Reference number: JC2143; MUID:94264407; PMID:7764866
A/Accession: JC2143
A/Molecule type: DNA
A/Residues: 1-1034 <MIC>
C/Comment: This protein consists of 52 repeats of closely related 16-amino acid motifs
C/Superfamily: ice nucleation protein
C/161-993/Region: R-domain

```

Search completed: August 2, 2004, 14:56:25
Job time : 10.4 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:36:12 ; Search time 5.3 Seconds
(without alignments)

1483.508 Million cell updates/sec

Title: US-09-543-407-26

Perfect score: 782

Sequence: 1 MKLLKVAFAAIVVGSALA.....DSSVMVRQVGFGNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	692	88.5	151	1	CSGA_SALTY
2	523	66.9	151	1	CSGA_ECOLI
3	501.5	64.1	152	1	CSGA_ECO57
4	112	14.3	151	1	CSGB_ECOLI
5	107.5	13.7	1322	1	ICEN_PANAN
6	107	13.7	151	1	CSGB_SALTY
7	107	13.7	151	1	CSGB_SALTY
8	106	13.6	1034	1	ICEN_PANAN
9	104.5	13.4	1258	1	ICEN_ERWHE
10	102.5	13.1	1567	1	ICEN_XANCT
11	102.5	13.1	1655	1	ONPB_RICCN
12	97.5	12.5	1028	1	OVG_DROME
13	96.5	12.3	1196	1	ICEV_PSEX
14	95.5	12.2	1656	1	ONPB_RICUA
15	94	12.0	599	1	GP63_LEICH
16	93	11.9	491	1	YK98_MYCTU
17	91.5	11.7	894	1	IFP3_HUMAN
18	91	11.6	678	1	YF48_MYCTU
19	90.5	11.6	1200	1	ICEN_PSEY
20	89.5	11.4	1317	1	N145_YEAST
21	88.5	11.3	646	1	GP63_LEIME
22	88.5	11.3	1148	1	ICEN_PSEX
23	88.5	11.3	1210	1	ICEN_PSEFL
24	87.5	11.2	602	1	GP63_LEINA
25	86.5	11.1	590	1	GP63_LEIDO
26	86	11.0	959	1	N100_YEAST
27	85	10.9	681	1	YDHE_SCHPO
28	84.5	10.8	548	1	CEAK_ECOLI
29	84.5	10.8	641	1	IMD_ARTGO
30	84.5	10.8	1300	1	120K_RICRI
31	84.5	10.8	1654	1	ONPB_RICRI
32	82.5	10.5	487	1	Y442_MYCTU
33	82	10.5	1327	1	TNKL_HUMAN

34 81.5 10.4 347 1 MSA2_PLAP2
35 81.5 10.4 1778 1 N189_SCHPO
36 81 10.4 365 1 ROAL_DROME
37 81 10.4 443 1 Y878_MYCTU
38 81 10.4 877 1 SEU_ARATH
39 80.5 10.3 1063 1 SPT5_YEAST
40 80 10.2 1293 1 MLE_DROME
41 79.5 10.2 1067 1 SGG_DROME
42 79 10.1 165 1 GRP1_ORVSA
43 79 10.1 346 1 RO21_XENLA
44 79 10.1 1302 1 FRPA_NEIMB
45 79 10.1 1645 1 ONPB_RICTY

Q03646 plasmodium
Q0utk4 schizosacch
P07909 drosophila
Q10540 mycobacteri
Q8w234 arabidopsis
P27692 saccharomyc
P24785 drosophila
P18431 drosophila
P25074 oryza sativ
P51989 xenopus lae
Q9X0k9 neisseria m
P96989 r outer mem

ALIGNMENTS

RESULT 1
ID CSGA_SALTY STANDARD; PRT; 151 AA.
AC P55225;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Major curlin subunit precursor (Fimbrin SEF17).
GN CSGA OR AGFA OR STM1144 OR STY1181 OR T1776.
OS Salmonella typhimurium,
OS Salmonella typhi, and
OS Salmonella enteritidis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602, 601, 592;
[1] _TaxID=602, 601, 592;
SEQUENCE FROM N.A.
RP SPECIES=S.typhimurium; STRAIN=SR-11;
RC MEDLINE=98117058; PubMed=9457880;
RA Romling U., Bian Z., Hammar M., Sierralta W.D., Normark S.;
RT "Curli fibers are highly conserved between Salmonella typhimurium and
RT Escherichia coli with respect to operon structure and regulation.";
RL J. Bacteriol. 180:722-731(1998).
[2]
SEQUENCE FROM N.A.
RP SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
RC MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
[3]
SEQUENCE FROM N.A.
RP SPECIES=S.typhi; STRAIN=CT18;
RC MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebatia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feitwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moulé S., O'Gaora P., Parry C.,
RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrrell B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
[4]
SEQUENCE FROM N.A.
RP SPECIES=S.typhi; STRAIN=Ty2 / ATCC 700931;
RC MEDLINE=22531367; PubMed=12644504;
RA Deng W., Liu S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodoyanni V., Schwartz D.C., Blattner F.R.;

"Comparative genomics of *Salmonella enterica* serovar Typhi strains Ty2 and CT18.";
 J. Bacteriol. 185:2330-2337(2003).
 [5] SEQUENCE FROM N.A.
 RC SPECIES=S.enteritidis; STRAIN=27655-3B;
 RX MEDLINE=96146512; PubMed=8550497;
 RA Collinson S.K., Clouthier S.C., Doran J.L., Baner P.A., Kay W.W.;
 RT "Salmonella enteritidis agfBAC operon encoding thin, aggregative
 fimbriae.";
 RL J. Bacteriol. 178:662-667(1996).
 RN [6] SEQUENCE OF 21-151 FROM N.A.
 RP SPECIES=S.enteritidis; STRAIN=27655-3B;
 RC SPECIES=S.enteritidis; STRAIN=27655-3B;
 RX MEDLINE=94013373; PubMed=8104955;
 RA Doran J.L., Collinson S.K., Burian J., Sarlos G., Todd E.C.D.,
 Munro C.K., Kay C.M., Baner P.A., Peterkin P.I., Kay W.W.;
 RT "DNA-based diagnostic tests for *Salmonella* species targeting agfA,
 the structural gene for thin, aggregative fimbriae.";
 RL J. Clin. Microbiol. 31:2263-2273(1993).
 RN [7] SEQUENCE OF 21-33.
 RP SPECIES=S.enteritidis; STRAIN=27655-3B;
 RC SPECIES=S.enteritidis; STRAIN=27655-3B;
 RX MEDLINE=91310586; PubMed=1677357;
 RA Collinson S.K., Emsedy L., Mueller K.-M., Trust T.J., Kay W.W.;
 RT "Purification and characterization of thin, aggregative fimbriae from
Salmonella enteritidis.";
 RL J. Bacteriol. 173:4773-4781(1991).
 CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 FIBRONECTIN.
 CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC
 DR EMBL; AJ002301; CA05317.1; -;
 DR EMBL; AE008749; AAL20074.1; -;
 DR EMBL; AL627269; CAD08268.1; -;
 DR EMBL; AE016840; AAO89399.1; -;
 DR EMBL; U43280; AAC43599.1; -;
 DR FIR; JC6039; JC6039.
 DR StyGene; SG10608; csGA.
 KW Fimbrin; Signal; Complete proteome.
 FT SIGNAL 1 20
 FT CHAIN 21 151 MAJOR CURLIN SUBUNIT.
 FT CONFLICT 134 151 SVMYRVQVFGNNATANY -> DSYTOVAS (IN
 REF. 6).
 FT SEQUENCE 151 AA; 15305 MW; B7DAC0D16B621359 CRC64;
 Query Match 88.5%; Score 692; DB 1; Length 151;
 Best Local Similarity 90.7%; Pred. No. 3.2e-51;
 Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;
 Qy 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNNHGGNNSGPDSTLSIYQVGSANAALYDQ 60
 Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNNHGGNNSGPDSTLSIYQVGSANAALQ 60
 Qy 61 LVTRVVTHEAHAGYNGADVGGADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGG 120
 Db 61 SDAREKSTTTQSGYNGADVGGADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGG 120
 Qy 121 NNAALVNOTASDSVMYRVQVFGNNATANY 151
 Db 121 NNAALVNOTASDSVMYRVQVFGNNATANY 151

RESULT 2
 CSGA_ECOLI STANDARD; PRT; 151 AA.
 ID CSGA_ECOLI
 AC P28307;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Major curlin subunit precursor.
 GN CSGA OR B1042.
 OS *Escherichia coli*.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; *Escherichia*.
 CX NCBI_TaxID=562;
 RN [1] SEQUENCE FROM N.A.
 RP STRAIN=K12 / W3110;
 RX MEDLINE=93211294; PubMed=8459772;
 RA Olsen A., Arngvist A.;
 RT "The RpoS sigma factor relieves H-NS-mediated transcriptional
 repression of csGA, the subunit gene of fibronectin-binding curli in
Escherichia coli.";
 RL Mol. Microbiol. 7:523-536(1993).
 RN [2] SEQUENCE FROM N.A.
 RP STRAIN=K12 / MC4100;
 RX MEDLINE=96414468; PubMed=8817489;
 RA Hammar M., Arngvist A., Bian Z., Olsen A., Normark S.;
 RT "Expression of two csG operons is required for production of
 fibronectin- and Congo red-binding curli polymers in *Escherichia coli*
 K-12.";
 RL Mol. Microbiol. 18:661-670(1995).
 RN [3] SEQUENCE FROM N.A.
 RP STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 Gregor J., Davis N.W., Kirpatrick H.A., Goeden M.A., Rose D.J.,
 Mau B., Shao Y.;
 RT "The complete genome sequence of *Escherichia coli* K-12.";
 RL Science 277:1453-1474(1997).
 RN [4] SEQUENCE FROM N.A.
 RP STRAIN=K12.
 RX MEDLINE=97061202; PubMed=8905232;
 RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
 Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
 Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
 Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
 Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 Yano M., Horiuchi T.;
 RT "A 718-kb DNA sequence of the *Escherichia coli* K-12 genome
 corresponding to the 12.7-28.0 min region on the linkage map.";
 RL DNA Res. 3:137-155(1996).
 RN [5] SEQUENCE OF 21-40.
 RP STRAIN=K12 / YWEL;
 RX MEDLINE=93028373; PubMed=1357528;
 RA Arngvist A., Olsen A., Pfeifer J., Russell D.G., Normark S.;
 RT "The Crl protein activates cryptic genes for curli formation and
 fibronectin binding in *Escherichia coli* HB101.";
 RL Mol. Microbiol. 6:2443-2452(1992).
 RN [6] SEQUENCE OF 21-31.
 RP MEDLINE=91310586; PubMed=1677357;
 RA Collinson S.K., Emsedy L., Trust T.J., Kay W.W.;
 RT "Purification and characterization of thin, aggregative fimbriae from
Salmonella enteritidis.";
 RL J. Bacteriol. 173:4773-4781(1991).
 CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 FIBRONECTIN.

```

CC -I- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L04979; AAA33616.1; --
CC DR EMBL; X90754; CAA62282.1; --
CC DR EMBL; A8000205; AAC74126.1; --
CC DR EMBL; D90741; BAA35832.1; --
CC DR EMBL; D90742; BAA35840.1; --
CC DR PIR; S70788; S70788.
CC DR EcoGene; EG11489; csGA.
CC Fimbrin; Signal; Complete proteome.
CC FT SIGNAL 1 20
CC FT CHAIN 21 151 MAJOR CURLIN SUBUNIT.
CC FT CONFLICT 151 AA; 15049 MW; C003470D208D595F CRC64;
CC FT SEQUENCE
CC
CC Query Match 66.9%; Score 523; DB 1; Length 151;
CC Best Local Similarity 68.9%; Pred. No. 4.1e-37;
CC Matches 104; Conservative 18; Mismatches 29; Indels 0; Gaps 0;
CC
CC QY 1 MKLLKVAAPAAIVVSGSALAGVGPWGCGGNGHGGNSGDPDSTLSIYQYGSNAALYDQ 60
CC | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
CC | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
CC Db 1 MKLLKVAIAAIVFSGSALAGVGPYQGGGNGHGGNSGNSPSENLYQYGGGNSALQ 60
CC
CC QY 61 LVTRVVTHEMAHAGYNGGADYVGQADNSTIELTQNGFRNNAITIDMNAKNSDITVGYGG 120
CC | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
CC | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
CC Db 61 TDARNSDLITITQHGCGGNGADYVGQSDSDSLDTQRFNGSATLQDNGKSEMTVKQFGG 120
CC
CC QY 121 NNAALVNQTASDSSVMYRVQVFGGNATANQY 151
CC | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
CC | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
CC Db 121 NGGAAVDQTASNSSVNVTVQVFGGNATAHQY 151
CC
CC RESULT 3
CC CSGA_ECO57 STANDARD; PRT; 152 AA.
CC ID ID CSGA_ECO57
CC AC Q83U24;
CC DT 28-FEB-2003 (Rel. 41, Created)
CC DT 28-FEB-2003 (Rel. 41, Last sequence update)
CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
CC DE Major curlin subunit precursor.
CC CS CSGA OR Z1676 OR EGS1420.
CC GN Escherichia coli O157:H7.
CC OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
CC OC Enterobacteriaceae; Escherichia.
CC OX NCBI_TaxID=83334;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
CC RX MEDLINE=21074935; PubMed=11205551;
CC RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
CC RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
CC RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
CC RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
CC RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
CC RA Welch R.A., Blattner F.R.;
CC RL "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
CC RN Nature 409:529-533(2001).

```

RT K-12.";
 RL Mol. Microbiol. 18:661-670(1995).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of *Escherichia coli* K-12.";
 RL Science 277:1453-1474 (1997).
 [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97061202; PubMed=8905232;
 RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
 RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horiuchi T.;
 RT "A 718-kb DNA sequence of the *Escherichia coli* K-12 genome
 RT corresponding to the 12.7-28.0 min region on the linkage map.";
 RL DNA Res. 3:137-155(1996).
 [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoudis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7.";
 RL Nature 409:529-533(2001).
 [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.,
 RA "Complete genome sequence of enterohaemorrhagic *Escherichia coli*
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22(2001).
 [6]
 RP SEQUENCE OF 1-21 FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=95157246; PubMed=7854117;
 RA Arqvist A., Olsen A., Normark S.;
 RT "Sigma S-dependent growth-phase induction of the *csgB* promoter in
 RT *Escherichia coli* can be achieved in vivo by sigma 70 in the absence
 RL of the nucleoid-associated protein H-NS.";
 RL Mol. Microbiol. 13:1021-1032(1994).
 CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
 CC CURLIN MONOMERS.
 CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X90754; CAA62281.1; -.

DR EMBL; AE000205; AAC74125.1; -.
 DR EMBL; D90741; BAA35831.1; -.
 DR EMBL; AE005315; AAG55787.1; -.
 DR EMBL; AP002554; BAB34842.1; -.
 DR PIR; C90806; C90806.
 DR PIR; G85665; G85665.
 DR PIR; S70787; S70787.
 DR EcoGene; EG12621; csgB.
 KW Fimbria; Signal; Complete proteome.
 FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 22 151 MINOR CURLIN SUBUNIT.
 SQ SEQUENCE 151 AA; 15882 MW; B18D266B9640:4B8 CRC64;
 Query Match 14.3%; Score 112; DB 1; Length 151;
 Best Local Similarity 29.8%; Pred. No. 0.0083;
 Matches 34; Conservative 17; Mismatches 49; Indels 14; Gaps 3;
 Qy 51 GSANAALYDQVLT--RVVTHEAHAGYNGADVGGQADNSTIELTQNGFR----- 98
 Db 17 GTAAAAGYDLANSEYNFAVNELSKSFNOAAIIGQAGTNNSAQLRQGGSKLLAVVAQEGS 76
 Qy 99 -NNATIDQWNAKNSDITVGYGNGNAALNQTASDSSVVMVGVGNNATANQY 151
 Db 77 SNRAKIDQTDYNL-AYIDQASANDASISQAYGNTAMIIQGGSGNKANITQY 129
 RESULT 5
 ICEA PANAN
 ID ICEA PANAN STANDARD; PRT; 1322 AA.
 AC P20469;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Ice nucleation protein inaa.
 GN INAA.
 OS Pantoea ananas (Erwinia uredovora).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Pantoea.
 OX NCBI_TaxID=553;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90092494; PubMed=2599095;
 RA Abe K., Watabe S., Emori Y., Watanabe M., Arai S.;
 RT "An ice nucleation active gene of *Erwinia ananas*. Sequence similarity
 RT to those of *Pseudomonas* species and regions required for ice
 RT nucleation activity.";
 RL FEBS Lett. 258:297-300(1989).
 CC -!- FUNCTION: Ice nucleation proteins enable bacteria to nucleate
 CC crystallization in supercooled water.
 CC -!- SUBCELLULAR LOCATION: Outer membrane (By similarity).
 CC -!- DOMAIN: CONTAINS MANY IMPERFECT REPEATS OF THE CONSENSUS
 CC OCTAPEPTIDE A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A
 CC REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.
 CC -!- SIMILARITY: Belongs to the bacterial ice nucleation protein
 CC family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X17316; CAA35194.1; -.
 DR PIR; S07053; S07053.
 DR EMBL; P06620; IINA.
 DR InterPro; IPR000258; Ice_nucleatn.
 DR Pfam; PF00818; Ice_nucleation; 69.
 DR PRINTS; PS00327; ICENUCLEATN.
 DR PROSITE; PS00314; ICE_NUCLEATION; 49.
 KW Ice nucleation; Repeat; Outer membrane.
 FT DOMAIN 162 1281 OCTAPEPTIDE PERIODICITY.

SQ SEQUENCE 1322 AA; 131094 MW; 89B0EE24AA837039 CRC64;
Query Match 13.7%; Score 107.5; DB 1; Length 1322;
Best Local Similarity 29.9%; Pred. No. 0.22;
Matches 38; Conservative 19; Mismatches 21; Indels 49; Gaps 8;
QY 34 GGNSSGPDSTLSYQYGSANAALYDQVTRVTHMAHAGYNGADVGQADNSTIELT 93
DB 933 GSTTAGPDSSL-IAGYGSTQGTAGNSILT-----AGYGS-----T 967
QY 94 QNGFRNATIDQWAKNSDITVGYG-----GNNAAALV-----NOTASDSSVMVRQVGF 143
DB 968 QTG-----QENSLLTGT-YGSTAGYESSLIAGVSTQTASFKSLM--AGY 1013
QY 144 NNATANQ 150
DB 1014 SSQTARE 1020
RESULT 6
CSGB_SALTY STANDARD; PRT; 151 AA.
ID CSGB_SALTY STANDARD; PRT; 151 AA.
AC Q827M3;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Minor curlin subunit precursor.
GN CSGB OR STY1180 OR T1777.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA Baker S., Baham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Doud L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Hogue A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrell B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.,
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RT and CT18.";
RN J. Bacteriol. 185:2330-2337(2003).
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
CC CURLIN MONOMERS.
CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
DB EMBL; AL627269; CAD09267.1; -;
DB EMBL; AE016840; AX069400.1; -;

KW Fimbria; Signal; Complete proteome.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 151 MINOR CURLIN SUBUNIT.
SQ SEQUENCE 151 AA; 16254 MW; 161C54326E573495 CRC64;
Query Match 13.7%; Score 107; DB 1; Length 151;
Best Local Similarity 30.4%; Pred. No. 0.22; Indels 16; Gaps 5;
Matches 35; Conservative 17; Mismatches 47;
QY 51 GSANAALYD--QLVTRVTHMAHAGYNGADVGQ--GADNST-----IELTQNGF 97
DB 17 GIATATNYDLARSEYNFAVNSLSKSFNQAAIIQGVGTDSARVRQEGSKLLSVISQEG 76
QY 98 RNATIDQWAKNSDIT-VGYGNNAAALVNTASDSSVMVRQVGFNNATANQY 151
DB 77 NNRKVDQ--AGNYNFAYIEOTGNANDASISOSAYGNSAAIIQKSGNKANITQY 129
RESULT 7
CSGB_SALTY STANDARD; PRT; 151 AA.
ID CSGB_SALTY STANDARD; PRT; 151 AA.
AC P55226;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Minor curlin subunit precursor (Fimbrin SEF17 minor subunit).
GN CSGB OR AGFB OR STM1143.
OS Salmonella typhimurium, and
OS Salmonella enteritidis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602, 592;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=SR-11;
RX MEDLINE=98117058; PubMed=9457880;
RA Romling U., Bian Z., Hammar M., Sierralta W.D., Normark S.;
RT "Curli fibers are highly conserved between Salmonella typhimurium and
RT Escherichia coli with respect to operon structure and regulation.";
RN J. Bacteriol. 180:722-731(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=JT2 / SGSC3412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=S.enteritidis; STRAIN=27655-3B;
RX MEDLINE=96146512; PubMed=8550497;
RA Collinson S.K., Clouthier S.C., Doran J.L., Baner P.A., Kay W.W.;
RT "Salmonella enteritidis agfBAC operon encoding thin, aggregative
RT fimbriae.";
RN J. Bacteriol. 178:662-667(1996).
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
CC CURLIN MONOMERS.
CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
DB EMBL; AL627269; CAD09267.1; -;
DB EMBL; AE016840; AX069400.1; -;

```

CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ002301; CAA05316.1; -.
CC EMBL; AB008749; AAL20073.1; -.
CC EMBL; U43280; AAC43598.1; -.
CC PIR; JC6040; JC6040.
CC StyGene; SG10609; csGB.
KW Fimbria; Signal; Complete proteome.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 151 MINOR CUELIN SUBUNIT.
SQ SEQUENCE 151 AA; 16182 MW; CFCF5430E6DD361D CRC64;

Query Match 13.7%; Score 107; DB 1; Length 151;
Best Local Similarity 30.4%; Pred. No. 0.022;
Matches 35; Conservative 17; Mismatches 47; Indels 16; Gaps 5;

QY 51 GSANAALYD-QLVTRVVTHEMAHAGYNGADVGO-GADNST-----IELTQNGP 97
DB 17 GIATATNYDLARSEYNFAVNLKSFNQAAIIQVGTDSARVRQEGSKLLSVISQEGG 76

QY 98 RNNATIDOWNAKNSDIT-VGQYGGNNAALVNQTASDSSVMVRQVGFNNATANQY 151
DB 77 NNRKVDQ--AGNYNFAVTEQGNANDASISQSAVGNAAIIQKSGNKANITQY 129

RESULT 8
ID ICEN PANAN STANDARD; PRT; 1034 AA.
AC Q47879;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ice nucleation protein inau.
GN INAU.
OS Pantoea ananas (Erwinia uredovora).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pantoea.
OX NCBI_TaxID=553;
RN [1]_TaxID=553;
RP SEQUENCE FROM N.A.
RC STRAIN=KUN-3;
RX MEDLINE=94264407; PubMed=7764866;
RA Michigami Y., Watabe S., Abe K., Obata H., Arai S.;
RT "Cloning and sequencing of an ice nucleation active gene of Erwinia
RT uredovora";
RL Biosci. Biotechnol. Biochem. 58:762-764(1994).
CC -!- FUNCTION: Ice nucleation proteins enable bacteria to nucleate
CC crystallization in supercooled water.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- DOMAIN: CONTAINS IMPERFECT REPEATS OF A CONSENSUS OCTAPEPTIDE
CC A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A REGIONAL 48-RESIDUE
CC PERIODICITY IS SUPERIMPOSED.
CC -!- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE
CC NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.
CC -!- SIMILARITY: Belongs to the bacterial ice nucleation protein
CC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D14992; JRA03636.1; -.
CC PIR; JC2143; JC2143.
CC HSP; P06620; IINA.
CC Pfam; PF00818; Ice_nucleatn.
CC PRINTS; PR00327; ICENUCLEATN.
CC PROSITE; PS00314; ICE_NUCLEATION; 34.
CC Ice nucleation; Repeat; Outer membrane.
KW

```

```

FT DOMAIN 162 993 OCTAPEPTIDE PERIODICITY.
SQ SEQUENCE 1034 AA; 103378 MW; FA222523D333EADD CRC64;

Query Match 13.6%; Score 106; DB 1; Length 1034;
Best Local Similarity 29.9%; Pred. No. 0.22;
Matches 40; Conservative 19; Mismatches 37; Indels 38; Gaps 9;

QY 27 GCGGNHNGGNSGPDSTLSIYQVGSANAALYDQLVTRVVTHEMAHAGYNGADVQCGAD 86
DB 531 GYGSTAGANSS-----LIAGYGTQTASVNSVLT-----AGYGTQTAREGSD 575

QY 87 NSTIELTQNGFRNNATIDOWNAKNSDITVQYGV-----GNNAALV-----NOTASDSSVM 136
DB 576 -----LTA-GYGTQTAE-----NSDLTTG-YGS-STAGYDSSLIAGYGTQTAGYHSIL 624

QY 137 VRQVGFNNATNQ 150
DB 625 T--AGYGTQTAE 636

RESULT 9
ID ICEN ERWHE STANDARD; PRT; 1258 AA.
AC P16239;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ice nucleation protein.
GN ICEE.
OS Erwinia herbicola.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pantoea.
OX NCBI_TaxID=549;
RN [1]_TaxID=549;
RP SEQUENCE FROM N.A.
RC STRAIN=ML;
RX MEDLINE=90152370; PubMed=2515997;
RA Warren G.J., Corotto L.V.;
RT "The consensus sequence of ice nucleation proteins from Erwinia
RT herbicola, Pseudomonas fluorescens and Pseudomonas syringae.";
RL Gene 85:239-242(1989).
CC -!- FUNCTION: Ice nucleation proteins enable bacteria to nucleate
CC crystallization in supercooled water.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- DOMAIN: CONTAINS 126 IMPERFECT REPEATS OF A CONSENSUS OCTAPEPTIDE
CC A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A REGIONAL 48-RESIDUE
CC PERIODICITY IS SUPERIMPOSED.
CC -!- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE
CC NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.
CC -!- SIMILARITY: Belongs to the bacterial ice nucleation protein
CC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M26382; AAA24823.1; -.
CC PIR; JQ0188; JQ0188.
CC HSP; P06620; IINA.
CC InterPro; IPR000258; Ice_nucleatn.
CC Pfam; PF00818; Ice_nucleatn; 65.
CC PRINTS; PR00327; ICENUCLEATN.
CC PROSITE; PS00314; ICE_NUCLEATION; 45.
CC Ice nucleation; Repeat; Outer membrane.
KW
FT DOMAIN 162 1217 OCTAPEPTIDE PERIODICITY.
SQ SEQUENCE 1258 AA; 125084 MW; 590E8A130077FBD4 CRC64;

Query Match 13.4%; Score 104.5; DB 1; Length 1258;
Best Local Similarity 29.5%; Pred. No. 0.36;

```



```
Matches 44; Conservative 23; Mismatches 37; Indels 45; Gaps 10;
QY 34 GGNSSGPDSTLSIYQYGSANAALYDQLVTRVVTHEMAHAGYN-----GA 79
Db 773 GSTTAGADSSL-IAGYGSTQTAGYHSILT-----AGYGSTQTAQERSDLTTCYGS 822
QY 80 DVGQADNSTIE---LTQN-GFRNNATI-----DQWNAKNSDITVQYQ-----GNNAAIV 126
Db 823 TSTAGADSSLIAGYGSTQTAGYHSILT-----AGYGSTQTAQERSDLTTCYGS 822
QY 127 -----NOTASDSSVMVRQVGFNNATANQ 150
Db 882 AGYGSTQTAGYNSILT--AGYGSTQTAQE 908

RESULT 10
ID ICEN_XANCT STANDARD; PRT; 1567 AA.
AC P18127;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ice nucleation protein.
GN INAX.
OS Xanthomonas campestris (pv. translucens).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=343;
[1]
RN RP SEQUENCE FROM N.A.
RC STRAIN=X565;
RX MEDLINE=91080859; PubMed=2259339;
RA Zhao J., Orser C.S.;
RT "Conserved repetition in the ice nucleation gene inax from
RT Xanthomonas campestris pv. translucens.";
RL Mol. Gen. Genet. 223:163-166(1990).
CC -!- FUNCTION: Ice nucleation proteins enable bacteria to nucleate
CC crystallization in supercooled water.
CC -!- SUBCELLULAR LOCATION: Outer membrane (By similarity).
CC -!- DOMAIN: CONTAINS 153 IMPERFECT REPEATS OF THE CONSENSUS
CC OCTAPEPTIDE A-G-Y-G-S-T-L-T; FURTHER ON A 16-RESIDUE AND A
CC REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.
CC -!- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE
CC NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.
CC -!- SIMILARITY: Belongs to the bacterial ice nucleation protein
CC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X52970; CAA37140.1; -.
CC HSSP; P06620; 1INA.
CC InterPro; IPR000258; Ice_nucleatn.
CC Pfam; PF00818; Ice_nucleation; 81.
CC PRINTS; PR00327; ICNUCLEATN.
CC PROSITE; PS00314; ICE_NUCLEATION; 57.
CC Ice_nucleation; Repeat; Outer membrane.
CC SEQUENCE 1567 AA; 152548 MW; C8B451D9595ECAD63 CRC64;

Query Match
Best Local Similarity 30.2%; Pred. No. 0.68;
Matches 45; Conservative 23; Mismatches 36; Indels 45; Gaps 11;
QY 34 GGNSSGPDSTLSIYQYGSANAALYDQLVTRVVTHEMAHAGYN-----GADV----- 81
Db 1055 GSTGTAGADSSL-IAGYGSTQTAGSDSSLT-----AGYGSTQTAQERSDLTTCYGS 1104
QY 82 --GQADNSTIE---LTQN-GFRNNATI-----DQWNAKNSDITVQYQ-----GNNAAIV 126
```

```
Db 1105 TGTAGADSSLIAGYGSTQTAGYSDSLNLTAGYGSTQTAQERSDLTTCYGS 1163
QY 127 -----NOTASDSSVMVRQVGFNNATANQ 150
Db 1164 AGYGSTQTAGYNSILT--TG YGSTQTAQE 1190

RESULT 11
ID OMPB_RICCN STANDARD; PRT; 1655 AA.
AC Q9KK3; O9KK98; O9KC45;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Outer membrane protein B precursor (168 kDa surface-layer protein)
DE (Surface protein antigen) (Cell surface antigen 5) (Scas) (rOmpB)
DE (rOmp B) [Contains: 120 kDa surface-exposed protein (Surface protein
DE antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide].
GN OMPB OR RC1085.
OS Rickettsia conorii.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=781;
[1]
RN RP SEQUENCE FROM N.A.
RC STRAIN=Malish 7;
RX MEDLINE=21442074; PubMed=11557893;
RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
RA Raoult D.;
RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
RL Science 293:2093-2098(2001).
[2]
RN RP SEQUENCE OF 33-1649 FROM N.A.
RC STRAIN=Indian tick typhus, and Malish 7;
RX MEDLINE=20393643; PubMed=10939649;
RA Roux V.; Raoult D.;
RT "Phylogenetic analysis of members of the genus Rickettsia using the
RT gene coding the outer-membrane protein rOmpB (ompB)."
RT Int. J. Syst. Evol. Microbiol. 50:1449-1455(2000).
[3]
RN RP SEQUENCE OF 353-1655 FROM N.A.
RC STRAIN=Malish 7;
RA Stenos J., Walker D.;
RT "The rickettsial outer membrane protein A and B genes of Rickettsia
RT australis, the most divergent rickettsia of the spotted fever group.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DDAT databases.
CC -!- FUNCTION: THE 120 kDa SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (By
CC similarity).
CC -!- FUNCTION: THE 32 kDa BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR
CC (By similarity).
CC -!- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-
CC layer with hexagonal symmetry (By similarity).
CC -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; A3008659; AAL03623.1; -.
CC EMBL; AF123721; AAF34124.1; -.
CC EMBL; AF123726; AAF34129.1; -.
CC EMBL; AF149110; AAD39533.1; -.
CC FIR; E97835; E37835.
CC InterPro; IPR006315; Autotransport.
CC InterPro; IPR005546; Autotransporter.
```



```

ID AC ICEV PSEX STANDARD; PRT; 1196 AA.
DE O33479;
DE 15-DEC-1998 (Rel. 37, Created)
DE 15-DEC-1998 (Rel. 37, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ice nucleation protein.
GN INAV.
OS Pseudomonas syringae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=317;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=INA5;
RX MEDLINE=97462815; PubMed=9323042;
RA Schmid D., Pridmore D., Capitani G., Battistutta R., Neeser J.-R.,
RA Jann A.
RT "Molecular organisation of the ice nucleation protein Inav from
RT Pseudomonas syringae."
RL FEBS Lett. 414:590-594 (1997).
CC -!- FUNCTION: Ice nucleation proteins enable bacteria to nucleate
CC crystallization in supercooled water.
CC -!- SUBUNIT: MEMBRANE ENVIRONMENT OR AGGREGATION SEEMS TO BE REQUIRED
CC FOR ICE NUCLEATION ACTIVITY.
CC -!- SUBCELLULAR LOCATION: Outer membrane (By similarity).
CC -!- DOMAIN: CONTAINS MANY IMPERFECT REPEATS OF THE CONSENSUS
CC OCTAPEPTIDE A-G-Y-G-S-T-L-T; FURTHER ON A 16-RESIDUE AND A
CC REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.
CC -!- SIMILARITY: Belongs to the bacterial ice nucleation protein
CC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AJ001086; CAA04521.1; -
CC HSP; P06620; LINA.
CC InterPro: IPR00258; Ice nucleatn.
CC Pfam: PF00819; Ice nucleation; 61.
CC PRINTS: PR00327; ICENUCLENTN.
CC PROSITE: PS00314; ICE NUCLEATION; 42.
CC Ice nucleation; Repeat; Outer membrane.
CC DOMAIN 172 1147 OCTAPEPTIDE PERIODICITY.
CC SEQUENCE 1196 AA; 117991 MW; C9E9974CB1731E68 CRC64;
CC
CC Query Match 12.3%; Score 96.5; DB 1; Length 1196;
CC Best Local Similarity 26.7%; Pred. No. 1.6;
CC Matches 43; Conservative 24; Mismatches 39; Indels 55; Gaps 11;
CC
CC 16 GSALAG-----VVPQWGG---GNHN-----GGNSSGPDSTLSIQYGSANAALYDQL 61
CC 175 GSTLSGDNNSRLIAGYGSNETAGNSDLIAGYGSTGTAGSDSL-VAGYGSTGTAGDGA 233
CC
CC 62 VTRVVTHEMAHAGYGN-----GADVGGADNSTLTQTNGFRNNATIDQWN 107
CC 234 LT-----AGVGSQTAREGSLNLTAGYGSTGTAGSDSLIA-----GYGSTQT---S 276
CC
CC 108 AKNSDITVGYGGNNAALVNTQSDSSVMVRQVFGNNATA 148
CC 277 GEDSLLTAG-YGS-----TQTAGEGSLNLT--AGYGSTGTA 308
CC
CC RESULT 14
CC OMPE RICJA STANDARD; PRT; 1656 AA.
ID -OMPB RICJA
AC O06653;
DE 30-MAY-2000 (Rel. 39, Created)
DE 30-MAY-2000 (Rel. 39, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE
DE Outer membrane protein B precursor (168 kDa surface-layer protein)
DE (Surface protein antigen) (Cell surface antigen 5) (Scas) (rOmpB)
DE (rOmp B) [Contains: 120 kDa surface-exposed protein (Surface protein
DE antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide].
GN OMPB.
OS Rickettsia japonica.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=35790;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YH;
RX Uchiyama T.;
RT "Sequencing of the gene encoding the protein rOmp B of Rickettsia
RT japonica."
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THE 120 kDa SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (By
CC similarity).
CC -!- FUNCTION: THE 32 kDa BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR
CC (By similarity).
CC -!- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-
CC layer with hexagonal symmetry.
CC -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB003681; BAA20138.1; -
CC InterPro: IPR006315; Autotransport.
CC InterPro: IPR005546; Autotransporter.
CC Pfam: PF03797; Autotransporter; 1.
CC TIGRFAMs: TIGR01414; autotrans_bar1; 2.
CC Antigen; S-layer; Cell wall.
CC CHAIN 1 1338 120 kDa SURFACE-EXPOSED PROTEIN.
CC FT CHAIN 1339 1656 32 kDa BETA PEPTIDE.
CC FT DOMAIN 528 533 POLY-GLY.
CC SEQUENCE 1656 AA; 168097 MW; 3132A69C9DD5999F CRC64;
CC
CC Query Match 12.2%; Score 95.5; DB 1; Length 1656;
CC Best Local Similarity 25.0%; Pred. No. 2.8;
CC Matches 36; Conservative 21; Mismatches 52; Indels 35; Gaps 7;
CC
CC 28 GGGNNGGNGSGSGPDSTLS-----IVQGSANAALYDQLVTR-----VVTHEMAHAGYGN 77
CC 1215 GGARNF--GTLGGPNFTVIGSNRFVNYGLIRAAQDYVITRTNNAENIVTNDITNSPFG 1272
CC
CC 78 GADVGGADNSTLTQTNGFRNNATIDQWNAKNSDITVGYGGNNAALVNTQSDSSVMV 137
CC 1273 APGVQGVN-TTFVNATNTAAVNNLL-----AKNS-----ADSNFVGTIVTDTSAI 1319
CC
CC 138 R-----QVGFNNATANQY 151
CC 1320 TNAQLDVAKD-QAQLGNRLGALRY 1343
CC
CC RESULT 15
CC GP63 LEICH STANDARD; PRT; 599 AA.
ID -GP63 LEICH
AC P15706;
DE 01-APR-1990 (Rel. 14, Created)
DE 01-APR-1990 (Rel. 14, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE Leishmanolysin precursor (EC 3.4.24.36) (Cell surface protease)
DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface
DE endopeptidase).
GN GP63.

```

OS Leishmania chagasi.
 CC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 OX NCBI_TaxID=44271;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=90205976; PubMed=2320059;
 RX Miller R.A., Reed S.G., Parsons M.;
 RT "Leishmania gp63 molecule implicated in cellular adhesion lacks an
 RL Arg-Gly-Asp sequence";
 RL Mol. Biochem. Parasitol. 39:267-274 (1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=92112918; PubMed=1370484;
 RX Ramamoorthy R., Donelson J.E., Paetz K.E., Maybodi M., Roberts S.C.,
 RA Wilson M.E.;
 RT "Three distinct RNAs for the surface protease gp63 are differentially
 RT expressed during development of Leishmania donovani chagasi
 RT promastigotes to an infectious form";
 RL J. Biol. Chem. 267:1888-1895 (1992).
 CC -!- FUNCTION: Has an integral role during the infection of macrophages
 CC in the mammalian host.
 CC -!- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and
 CC P1' and basic residues at P2 and P3'. A model nonapeptide is
 CC cleaved at Ala-Tyr-Leu-Lys-Lys-
 CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
 CC -!- SIMILARITY: Belongs to peptidase family M8.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M80672; AA29238.1; -;
 DR EMBL; M28527; AA29235.1; -;
 DR PIR; A44951; A44951.
 DR HSP; P08148; LML.
 DR MEROPS; M08.001; -;
 DR InterPro; IPR006025; Pept M Zn BS.
 DR InterPro; IPR001577; Peptidase M8.
 DR Pfam; PF01457; Peptidase M8.1.
 DR PRINTS; PR00782; LSHMANOLYSIN
 DR PROSITE; PS00142; ZINC_PROTEASE; 1.
 KW Hydrolase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;
 KW Zymogen; Signal; Cell adhesion; GPI-anchor; Lipoprotein.
 FT SIGNAL 1 39
 FT PROPEP 40 97
 FT CHAIN 98 574
 FT PROPEP 575 599
 FT METAL 261 261
 FT ACT_SITE 262 262
 FT METAL 265 265
 FT METAL 331 331
 FT METAL 122 139
 FT DISULFID 122 139
 FT DISULFID 138 227
 FT DISULFID 311 393
 FT DISULFID 390 452
 FT DISULFID 403 422
 FT DISULFID 412 486
 FT DISULFID 463 507
 FT DISULFID 512 562
 FT DISULFID 532 555
 FT CARBOHYD 237 297
 FT CARBOHYD 394 394
 FT LIPID 574 574
 FT SEQUENCE 599 AA; 63848 NW; 746730AE8E2A2E7C CRC64;
 SQ
 Query Match 12.0%; Score 94; DB 1; Length 599;
 Best Local Similarity 54.8%; Pred. No. 1.2;

Matches 23; Conservative 5; Mismatches 8; Indels 6; Gaps 2;
 QY 52 SAN-AALYDQLVTRVVTTHMAHA-----GYNGADVGQADN 87
 Db 244 ANIASRYDQLVTRVVTTHMAHALGFSVGFEGARILEISN 285
 Search completed: August 2, 2004, 14:49:32
 Job time : 6.3 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:39:17 ; Search time 29.7 Seconds

(without alignments)

1604.150 Million cell updates/sec

Title: US-09-543-407-26

Perfect score: 782

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVGFGNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 31518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL 25:*

- 1: sp_archea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriap:*
- 17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	683	87.3	152	033802	Q33802 salmonella
2	580.5	74.2	150	07X243	Q7x243 citrobacter
3	532	68.0	149	07X240	Q7x240 citrobacter
4	508.5	65.0	152	Q8CW63	Q8cw63 escherichia
5	442.5	56.6	150	07X237	Q7x237 enterobacte
6	304	38.9	76	Q54069	Q54069 salmonella
7	122	15.6	29	Q9S3U5	Q9s3j5 escherichia
8	121	15.5	502	16 Q8EIH4	Q8eih4 shewanella
9	114.5	14.6	145	16 Q8UGN9	Q8ueg9 agrobacteri
10	113	14.5	2174	16 Q92UJ8	Q92uu8 rhizobium m
11	112	14.3	151	16 Q7UCZ1	Q7ucz1 shigella fl
12	112	14.3	160	16 Q8CW64	Q8cw64 escherichia
13	112	14.3	160	16 Q83RU7	Q83ru7 shigella fl
14	110	14.1	151	2 Q7X244	Q7x244 citrobacter
15	108.5	13.9	1333	16 Q8PD38	Q8pd38 xanthomonas
16	107	13.7	1422	16 Q8EFU3	Q8efu3 shewanella

17	105.5	13.5	645	16	Q7U1C5	Q7u1c5 mycobacteri
18	105.5	13.5	646	16	O53818	O53818 mycobacteri
19	105	13.4	589	16	Q7TX53	Q7tx53 mycobacteri
20	105	13.4	590	16	O53309	O53309 mycobacteri
21	105	13.4	603	16	Q8VU66	Q8vj66 mycobacteri
22	104.5	13.4	151	2	Q7X238	Q7x238 enterobacte
23	104	13.3	490	16	Q8EYJ9	Q8eyj9 leptospira
24	103.5	13.2	91	2	Q9S3J8	Q9s3j8 escherichia
25	103.5	13.2	1306	2	Q93N36	Q93n36 pantoea ana
26	102.5	13.1	552	16	P96840	P96840 mycobacteri
27	102.5	13.1	552	16	Q7TW76	Q7tw76 mycobacteri
28	102.5	13.1	623	16	Q8VIF0	Q8viy0 mycobacteri
29	102.5	13.1	1616	2	Q9KKA1	Q9kkal rickettsia
30	102	13.0	130	16	Q89J14	Q89ji4 bradyrhizob
31	102	13.0	139	16	Q8EIH3	Q8eih3 shewanella
32	101.5	13.0	152	2	Q7X241	Q7x241 citrobacter
33	101.5	13.0	1616	2	Q9KKA5	Q9kka5 rickettsia
34	101	12.9	157	16	Q88H30	Q88hg0 pseudomonas
35	100.5	12.9	1616	2	Q9KKA9	Q9kka9 rickettsia
36	100	12.8	191	10	Q7XDR3	Q7xdr3 oryza sativ
37	100	12.8	3275	16	Q8VKM3	Q8vkm3 mycobacteri
38	100	12.8	3300	16	O06304	O06304 mycobacteri
39	100	12.8	3507	16	Q7U270	Q7u270 mycobacteri
40	99.5	12.7	1613	2	Q84OU5	Q84ou5 rickettsia
41	98.5	12.6	154	16	Q89J15	Q89jis bradyrhizob
42	98.5	12.6	453	5	Q9N6M6	Q9nem6 drosophila
43	98.5	12.6	738	5	O02402	O02402 pinctada fu
44	98.5	12.6	1286	2	Q84IY5	Q84iy5 campylobact
45	98.5	12.6	3659	16	Q98LN6	Q98ln6 rhizobium 1

ALIGNMENTS

RESULT 1

O33802 PRELIMINARY; PRT; 152 AA.

AC O33802;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE AgfA protein (Fragment).
GN AGFA.

OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;

RN [1]_SEQUENCE FROM N.A.

RP MEDLINE=98053391; PubMed=9393832;

RA Sukupolvi S.S., Lorentz R.G., Gordon J.I., Bian Z., Pfeifer J.D.,

RA Normark S.J., Rhen M.;

RT "Expression of thin, aggregative fimbriae promotes interaction of

RT Salmonella typhimurium SR-11 with mouse small intestinal epithelial

RT cells";

RL Infect Immun. 65:5320-5325(1997).

DR EMBL; AJ000514; CAA04151.1; -.

FT NON_TER 152 152

SQ SEQUENCE 152 AA; 15401 MW; 9DA7DADC2364B006 CRC64;

Query Match 87.3%; Score 683; DB 2; Length 152;

Best Local Similarity 89.4%; Pred. No. 2.8e-48;

Matches 135; Conservative 3; Mismatches 13; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNNHNGSGSPDSTLSITVYGSANNAALYDQ 60

DB 1 MKLLKVAFAAIVVSGSAVAGVVPQWGGGNNHNGSGSPDSTLSITVYGSANNAALQ 60

QY 61 LVTRVVTTHMAHAGYNGADVQGGADNSTIELTQNGFRNNATIDOWNAKNSDITVQYGG 120

DB 61 SPARKSETITQSGYNGADVQGGADNSTIELTQNGFRNNATIDOWNAKNSDITVQYGG 120

QY 121 NNAALVNQTSADSSVMVRQVGFGNATANQY 151

RT	"Production of Cellulose and Curli Fimbriae by Members of the Family
RL	Enterobacteriaceae isolated from the Human Gastrointestinal tract.";
RL	Infect. Immun. 72:4151-4158 (2003).
DR	EMBL; AJ515702; CAD56678.1; -.
SEQ	SEQUENCE 150 AA; 15112 MW; 5D8BB2D872DF15F3 CRC64;
Query Match	
Best Local Similarity 56.6%; Score 442.5; DB 2; Length 150;	
Matches 91; Conservative 25; Mismatches 34; Indels 1; Gaps 1;	
QY	1 MKLLKVAAPAAIYVSGSALAGVWPQWGGGNNHGGGNSGPDSTLSIYQYGSANAALYDQ 60
Db	1 MKFIKVAALAAIYVSGSAGMAGMIHQ -GWGHGHGHGGYGGPNSTLIYQNGGNSALALQ 59
QY	61 LVTRVVTTHMAHAGYNGADYGCAGADNSTIELTQNGFRNNATIDQWNAKNSDIITVGYGG 120
Db	60 TDARNSVLNISTGGGGADYGCQSDSSINLIQNGFGNSATLDQWNSKDSVMVVSQYGG 119
QY	121 NNAALVNQTASDSSVMVRQVGFNNATANQV 151
Db	120 LNALVDQATASNTVNTVQIGFGNHATAHQY 150
RESULT 6	
Q54069	PRELIMINARY; PRT; 76 AA.
ID	Q54069
AC	Q54069;
DT	01-NOV-1996 (TrEMBLrel. 01, Created)
DT	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE	SEF17 fimbriin (Fragment).
GN	AGFA.
OS	Salmonella enteritidis.
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC	Enterobacteriaceae; Salmonella.
OX	NCBI_TaxID=592;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=SE30;
RA	Cox J.M., Eglezos S., Woolcock J.B.;
RT	"Virulence of Salmonella enteritidis in chickens correlates with
RT	colony morphology and expression of SEF17 fimbriae.";
RL	Submitted (Apr-1996) to the EMBL/Genbank/DBJ databases.
DR	EMBL; U53207; AAA98671.1; -.
FT	NON TER 1 1
FT	NON TER 76 76
SEQ	SEQUENCE 76 AA; 7704 MW; 2FD5411241A7BCB1 CRC64;
Query Match	
Best Local Similarity 38.9%; Score 304; DB 2; Length 76;	
Matches 61; Conservative 2; Mismatches 13; Indels 0; Gaps 0;	
QY	30 GNHGGGNSGPDSTLSIYQYGSANAALYDQLVTRVVTTHMAHAGYNGADYGCAGADNST 89
Db	1 GNHGGGNSGPDSTLSIYQYGSANAALQSDARKSETTITQSYGNGADYGCAGADNST 60
QY	90 IELTQNGFRNNATIDQ 105
Db	61 IELTQNGFRNNATIDQ 76
RESULT 7	
Q9S3J5	PRELIMINARY; PRT; 29 AA.
ID	Q9S3J5
AC	Q9S3J5;
DT	01-MAY-2000 (TrEMBLrel. 13, Created)
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT	01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE	Curlin subunit monomer (Fragment).
GN	CSGA.
OS	Escherichia coli.
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC	Enterobacteriaceae; Escherichia.

```

Q8U6N9
ID Q8U6N9 PRELIMINARY; PRT; 145 AA.
AC Q8U6N9;
DT 01-JUN-2002 (TRENBLrel. 21, Created)
DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Hypothetical protein Atu4768.
GN ATU4768 OR AGR L.228.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Strubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.P. Jr., Woo L.,
RA Chen Y., Faulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutyavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
RA Raymond C., Rouse G., Saenphimachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Lingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nestor E.W.;
RA "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58.";
RL Science 294:2317-2323(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Qurillo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houlam C., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA Wollam K., Allinger M., Doughy D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gursion J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RA "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58.";
RL Science 294:2323-2328(2001).
DR EMBL; AE009405; AAL45562.1; -.
DR EMBL; AE008209; AAK38682.1; -.
DR PIR; AD3143; AD3143.
DR PIR; H98144; H98144.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 145 AA; 14984 MW; DEDC870E1713D51A CRC64;

Query Match 14.6%; Score 114.5; DB 16; Length 145;
Best Local Similarity 25.2%; Pred. No. 0.064;
Matches 39; Conservative 26; Mismatches 59; Indels 31; Gaps 4;

QY 3 LLKVAFAAIVVSGSALAGVVPQWG-----CGNNGGNSGSPDSTLSIYQYGSANA 55
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1 MIRKSFIAISALVALVGLSAAAPAVANDVRIDYQYGWSNAGAGQYGNRTYQNGYIN- 59
QY 56 ALYDQLVPRVVTTHMAHAGYNGADVGGADNNTIELTQNGFRNNAIDQWNAKNSDITV 115
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 60 -----RIVGHQY---GRNLSAVGQEGHDNYGSTTQNGNRNVAGI----- 96
QY 116 GOYGGNNAA--LVNQTSADSSVWVRQVG 150
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 97 GQFGSNHTTILTDQNGNIAAGVQVGRGCSANVSQ 131

RESULT 10
Q92U08
ID Q92U08 PRELIMINARY; PRT; 2174 AA.
AC Q92U08;
DT 01-DEC-2001 (TRENBLrel. 19, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Hypothetical glycine-rich protein Smb21548.
GN RB0989 OR Smb21548.

Q8U6N9
OS Rhizobium meliloti (Sinorhizobium meliloti).
OG Plasmid pSymB (megaplasmid 2).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=1021;
RC MEDLINE=21396508; PubMed=11481431;
RA Finan T.M., Weidner S., Wong K., Buhrmester J., Chain P.,
RA Vornolter F.J., Hernandez-Lucas I., Becker A., Cowie J.,
RA Golding B., Puchler A.;
RA "The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-
RT fixing endosymbiont Sinorhizobium meliloti.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).
DR EMBL; AL603645; CAC49389.1; -.
DR PIR; E95865; E95865.
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
DR GO; GO:0004601; F:peroxidase activity; IEA.
DR GO; GO:0006979; P:response to oxidative stress; IEA.
DR InterPro; IPR005546; Autotransporter.
DR InterPro; IPR002016; Peroxidase.
DR InterPro; IPR002173; PfkB.
DR Pfam; PF03797; Autotransporter; 1.
DR PROSITE; PS00435; PEROXIDASE_1; 1.
DR PROSITE; PS00583; PFkB_KINASES_1; 2.
KW Plasmid; Hypothetical protein; Complete proteome.
SQ SEQUENCE 2174 AA; 203314 MW; 008BB68297B44182 CRC64;

Query Match 14.5%; Score 113; DB 16; Length 2174;
Best Local Similarity 27.0%; Pred. No. 2.3;
Matches 40; Conservative 21; Mismatches 51; Indels 36; Gaps 7;

QY 11 AIVSGSALAGVVPQ--WGGGNGHNGGNSGSPDSTLSIYQYGSANAALYDQLVPRVVT 68
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 693 AIATAGAGAGVILAQSIGGGN---GGNATGGDAGFGSGFTGGGGG----- 737
QY 69 EMAHAGYNGADVGG-----QGADNSTI--ELTQNGFRNNAIDQWNAK--NSDITV 115
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 738 ----GGYANTANVGFKGLTLTQGSRAAGIVAQSVGGGGTGTASSISAGIGFTASVAV 793
QY 116 GOYGGNNAA--LVNQTSADSSVWVRQVG 141
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 794 GGTGGNGGAGGEVSLSLTDLSAIRTQGGG 821

RESULT 11
Q7UCZ1
ID Q7UCZ1 PRELIMINARY; PRT; 151 AA.
AC Q7UCZ1;
DT 01-OCT-2003 (TRENBLrel. 25, Created)
DT 01-OCT-2003 (TRENBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Minor curlin subunit.
GN CSGB OR S1108.
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2457T / ATCC 700930 / Serotype 2a;
RX MEDLINE=22590274; PubMed=12704152;
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
RA Schwartz D.C., Blattner F.R.;
RA "Complete genome sequence and comparative genomics of Shigella
RT flexneri serotype 2a strain 2457T.";
RL Infect. Immun. 71:2775-2786(2003).
DR EMBL; AB016981; AAP16542.1; -.
SQ SEQUENCE 151 AA; 15868 MW; 5D5D266B964014A0 CRC64;

```



```
Query Match 14.3%; Score 112; DB 16; Length 151;
Best Local Similarity 29.8%; Pred. No. 0.11;
Matches 34; Conservative 17; Mismatches 49; Indels 14; Gaps 3;

QY 51 GSANAALYDQVLT--RVVTHEMAHAGYNGADVGQADNSTIETQNGFR----- 98
DB 17 GIAAAGYDLANSEYNFAVNELSKSFNQAAIIGQAGTNSAQLRQGGSKLLAVVAQEGS 76

QY 99 -NNATIDOWNAKNSDITVGQYGGNNAALVNOTASDSSVMVRQVGFNNATANQY 151
DB 77 SNRAKIDQTDGYNL-AVIDQAGSANDASISQAGYNTAMIIQKSGNKANITQY 129

RESULT 12
Q8CW64 PRELIMINARY; PRT; 160 AA.
AC Q8CW64;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Minor curlin subunit precursor.
DE CSGB OR C1305.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.P., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.; Perna N.T.,
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).
DR EMBL: AE016759; AAN79778.1; -.
KW Complete proteome.
SQ SEQUENCE 160 AA; 16963 MW; 49F68448D979B986 CRC64;

Query Match 14.3%; Score 112; DB 16; Length 160;
Best Local Similarity 29.8%; Pred. No. 0.12;
Matches 34; Conservative 17; Mismatches 49; Indels 14; Gaps 3;

QY 51 GSANAALYDQVLT--RVVTHEMAHAGYNGADVGQADNSTIETQNGFR----- 98
DB 26 GIAAAGYDLANSEYNFAVNELSKSFNQAAIIGQAGTNSAQLRQGGSKLLTVVAQEGS 85

QY 99 -NNATIDOWNAKNSDITVGQYGGNNAALVNOTASDSSVMVRQVGFNNATANQY 151
DB 86 SNRAKIDQTDGYNL-AVIDQAGSANDASISQAGYNTAMIIQKSGNKANITQY 138

RESULT 13
Q83R7 PRELIMINARY; PRT; 160 AA.
AC Q83R7;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Minor curlin subunit precursor, similar to CsgA.
DE CSGB OR SF1035.
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=301 / Serotype 2a;
RX MEDLINE=2272406; PubMed=12384590;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
```

```
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding X., Chen S.,
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Yu J.;
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RT through comparison with genomes of Escherichia coli K12 and O157.";
RL Nucleic Acids Res. 30:4432-4441 (2002).
DR EMBL: AE015131; AAN42658.1; -.
KW Complete proteome.
SQ SEQUENCE 160 AA; 16919 MW; 50269F5268D2A32F CRC64;

Query Match 14.3%; Score 112; DB 16; Length 160;
Best Local Similarity 29.8%; Pred. No. 0.12;
Matches 34; Conservative 17; Mismatches 49; Indels 14; Gaps 3;

QY 51 GSANAALYDQVLT--RVVTHEMAHAGYNGADVGQADNSTIETQNGFR----- 98
DB 26 GIAAAGYDLANSEYNFAVNELSKSFNQAAIIGQAGTNSAQLRQGGSKLLAVVAQEGS 85

QY 99 -NNATIDOWNAKNSDITVGQYGGNNAALVNOTASDSSVMVRQVGFNNATANQY 151
DB 86 SNRAKIDQTDGYNL-AVIDQAGSANDASISQAGYNTAMIIQKSGNKANITQY 138

RESULT 14
Q7X244 PRELIMINARY; PRT; 151 AA.
AC Q7X244;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Nucleation component of curlin monomers.
DE CSGB.
OS Citrobacter sp. Fec2.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Citrobacter.
OX NCBI_TaxID=213763;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fec2;
RA Zogaj X., Bokranz W., Nimitz M., Romling U.;
RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RT Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158 (2003).
DR EMBL: AJ515700; CAD56671.1; -.
KW Complete proteome.
SQ SEQUENCE 151 AA; 16158 MW; BD00AF57E1400704 CRC64;

Query Match 14.1%; Score 110; DB 2; Length 151;
Best Local Similarity 27.2%; Pred. No. 0.16;
Matches 31; Conservative 22; Mismatches 47; Indels 14; Gaps 3;

QY 51 GSANAALYDQVLT--RVVTHEMAHAGYNGADVGQADNSTIETQNGFR----- 98
DB 17 GIASATSYDLAHSEYNFAVNELSKSFNQAAIIGQAGTNSAKMRQEGSKLLSVVSEQG 76

QY 99 -NNATIDOWNAKNSDITVGQYGGNNAALVNOTASDSSVMVRQVGFNNATANQY 151
DB 77 SNRAKVDQSGAYNF-AYIAQSGHSNDASISQSNYGNNTAMIIQKSGNKANITQY 129

RESULT 15
Q8PD38 PRELIMINARY; PRT; 1333 AA.
AC Q8PD38;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Ice nucleation protein.
DE XCC0507.
GN Xanthomonas campestris (pv. campestris).
OS Xanthomonas campestris (pv. campestris).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=340;
RN [1]
```

```

RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33913 / NCPPB 528;
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.B.A.,
RA Camarotte G., Canavan F., Cardozo J., Chambergo F., Ciapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Fornighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locati E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spindola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White P.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities."
RL Nature 417:453-463(2002).
DR EMBL; AE012148; AAM39823.1; -.
DR GO; GO:0009279; C:external outer membrane (sensu Gram-negativ. . .; IEA.
DR InterPro; IPR000258; Ice_nucleatn.
DR Pfam; PF00818; Ice_nucleation; 68.
DR PRINTS; PR00327; ICENUCLEATN.
DR PROSITE; PS00314; ICE_NUCLEATION; 40.
KW Complete proteome.
SQ SEQUENCE 1333 AA; 131306 MW; 326078458D0E4842 CRC64;

```

Query Match 13.9%; Score 108.5; DB 16; Length 1333;
Best Local Similarity 26.3%; Fred. No. 3;
Matches 46; Conservative 22; Mismatches 48; Indels 59; Gaps 9;

```

QY 5 KVAFAAIIWVG-----SALAGVWPQGGGNGGNGSGPDSLTSLIYQYGS 53
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : :
856 QTAGYKSLITGYGSTQTAQSSSLIAGY-----GSSMAGPSSSL-IAGYST 903
QY 54 NAALYDQIVTRVVTHEMAHAGYNGADVGGADNSTIELTQNGFRNNATID----- 104
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : :
904 QTAGYDSFLT-----AGYGS-----TQTAQSSSLITGYGSTSTAFQSSLIAGYS 950
QY 105 -QWNAKNSDITVGOYGGNNAL-----VNOTASDSVMVQVQFGNNATA 148
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : :
951 TQTAGYESTLTAG-YGSTQTQAEISWLTGYGSTQTAGHGSILT--AGYGSNSTA 1002

```

Search completed: August 2, 2004, 14:54:41
Job time : 30.7 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:35:42 ; Search time 44.9 Seconds

(without alignments)
950.215 Million cell updates/sec

Title: US-09-543-407-28

Perfect score: 775

Sequence: 1 MKLLKVAAPRAIVVSGSALA.....DSSVMVRQVFGNNATANQV 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*

- 1: Geneseq1980s:*
- 2: Geneseq1990s:*
- 3: Geneseq2000s:*
- 4: Geneseq2001s:*
- 5: Geneseq2002s:*
- 6: Geneseq2003as:*
- 7: Geneseq2003bs:*
- 8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	775	100.0	151	3 AAB36354	Aab36354 Agfa::PT3
2	712	91.9	151	3 AAB36351	Aab36351 Agfa::PT3
3	683	88.1	151	2 AAR74625	Aar74625 Agfa sequ
4	683	88.1	151	3 AAB36341	Aab36341 Salmonell
5	678	87.5	151	2 AAW23570	Aaw23570 Salmonell
6	662	85.4	151	3 AAB36350	Aab36350 Agfa::PT3
7	617	79.6	151	3 AAB36353	Aab36353 Agfa::PT3
8	613	79.1	151	3 AAB36355	Aab36355 Agfa::PT3
9	605	78.1	151	3 AAB36346	Aab36346 Agfa::PT3
10	603	77.8	151	3 AAB36347	Aab36347 Agfa::PT3
11	601	77.5	151	3 AAB36349	Aab36349 Agfa::PT3
12	600	77.4	151	3 AAB36352	Aab36352 Agfa::PT3
13	568	73.3	151	3 AAB36348	Aab36348 Agfa::PT3
14	509	65.7	151	3 AAB36343	Aab36343 Escherich
15	504	65.0	151	7 ABR8651	Abr8651 E. coli C
16	498	64.3	120	2 AAR62761	Aar62761 Agfa sequ
17	498	64.3	120	2 AAW23569	Aaw23569 Salmonell
18	435	55.1	142	2 AAR52664	Aar52664 Fibronect
19	359	46.3	122	2 AAR52663	Aar52663 FNB curli
20	147	19.0	45	3 AAB36316	Aab36316 Salmonell
21	132	17.0	22	3 AAB36318	Aab36318 Salmonell
22	113	14.6	24	7 ABR82644	Abr82644 E. coli c
23	111	14.3	22	3 AAB36322	Aab36322 Salmonell
24	111	14.3	22	3 AAB36327	Aab36327 Salmonell
25	111	14.3	22	3 AAB36337	Aab36337 Salmonell

ALIGNMENTS

RESULT 1

AAB36354
ID AAB36354 standard; protein; 151 AA.
XX
AC AAB36354;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa::PT3#9 amino acid sequence SEQ ID NO:28.
XX
KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO2000060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-).UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR N-PSDB; AAC64630.
XX
PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 138; 139pp; English.
XX
CC The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAf) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native

26 109 14.1 23 3 AAB36340 Salmonell
27 109 14.1 23 3 AAB36324 Salmonell
28 109 14.1 23 3 AAB36319 Salmonell
29 102 13.2 26 7 ABR82649 E. coli V
30 96 12.4 19 3 AAB36323 Salmonell
31 96 12.4 19 3 AAB36336 Salmonell
32 96 12.4 19 3 AAB36328 Salmonell
33 94.5 12.2 151 3 AAB36342 Salmonell
34 92 11.9 1074 6 ABU22692 Protein e
35 91.5 11.8 502 2 ABR82642 E. coli N
36 90 11.6 24 7 ABR82647 E. coli c
37 89.5 11.5 151 3 ABR82647 E. coli c
38 89 11.5 24 7 ABR82647 E. coli c
39 89 11.5 354 7 ABQ23520 Mycobacte
40 87.5 11.3 249 3 AAY69523 Anti-CD38
41 87.5 11.3 447 3 AAG29728 Arabidops
42 87.5 11.3 468 3 AAG29727 Arabidops
43 87.5 11.3 842 5 ABP66189 Bifidobac
44 87.5 11.3 1419 5 ABP69842 Human pol
45 87.5 11.3 1477 5 ABP69841 Human pol

CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbriae protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;

Query Match 100.0%; Score 775; DB 3; Length 151;
 Best Local Similarity 100.0%; Pred. No. 2.5e-67;
 Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSGDPSTLSIYQYGSANAALALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSGDPSTLSIYQYGSANAALALQ 60
 QY 61 SDARKSETTITQSGYNGADVDQLVTRVVTTHMAHAFNNATIDQWNAKNSDIIVGQYGG 120
 DB 61 SDARKSETTITQSGYNGADVDQLVTRVVTTHMAHAFNNATIDQWNAKNSDIIVGQYGG 120
 QY 121 NNAALVNQTASDSSVMVQVFGNNATANQY 151
 DB 121 NNAALVNQTASDSSVMVQVFGNNATANQY 151

RESULT 2
 AAB36351
 ID AAB36351 standard; protein; 151 AA.
 AC AAB36351;
 DT 26-FEB-2001 (first entry)
 DE Agfa::PT3#6 amino acid sequence SEQ ID NO:22.
 KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 PN WO200060102-A2.
 XX 12-OCT-2000.
 XX 05-APR-2000; 2000WO-CA000356.
 XX 05-APR-1999; 99US-0127888P.
 XX (UYVI-) UNIV VICTORIA.
 XX White AP, Doran JL, Collison SK, Kay WW;
 XX WPI; 2000-672631/65.
 XX N-PSDB; AAC64627.
 XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 XX which encodes foreign epitope or antigen, expresses recombinant Agfa
 XX protein useful for eliciting immune response in animal.
 XX Disclosure; Page 137; 139pp; English.

XX The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEB17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbriae protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;

Query Match 91.9%; Score 712; DB 3; Length 151;
 Best Local Similarity 91.1%; Pred. No. 3.3e-61;
 Matches 144; Conservative 0; Mismatches 0; Indels 14; Gaps 2;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSGDPSTLSIYQYGSANAALALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSGDPSTLSIYQYGSANAALALQ 60
 QY 61 SDARKSETTITQSGYNGADVDQLVTRVVTTHMAHAFNNATIDQWNAKNSDI 113
 DB 61 SDARKSETTITQSGYNGADVDQLVTRVVTTHMAHAFNNATIDQWNAKNSDI 113
 QY 114 TVGQYGGNNAALVNQTASDSSVMVQVFGNNATANQY 151
 DB 114 TVGQYGGNNAALVNQTASDSSVMVQVFGNNATANQY 151

RESULT 3
 AAR74625
 ID AAR74625 standard; protein; 151 AA.
 AC AAR74625;
 XX 25-MAR-2003 (revised)
 DT 26-JUN-1995 (first entry)
 DE Agfa sequence.
 DE Salmonella; Agfa; vaccine.
 KW Salmonella.
 OS Salmonella.
 XX WO9425598-A2.
 XX 10-NOV-1994.
 XX 26-APR-1994; 94WO-IB000207.
 XX 26-APR-1993; 93US-00054452.
 XX (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
 XX (KING/) KING J.
 XX

```

PI Kay WW, Collinson SK, Clouthier SC, Doran JL;
XX WPI: 1994-358275/44.
DR N-PSDB; AAC67467.
XX
XX Elliciting an immune response to Salmonella - using attenuated Salmonella
PT strains, vector constructs, or compsns. contg. fimbrial type proteins.
XX
XX Disclosure; Fig 7B; 95pp; English.
XX
XX The Salmonella AgfA protein and DNA are used in vaccine and genetic
CC immunization compositions, respectively, to elicit an immune response to
CC Salmonella in animals (e.g. food producing animals) and humans. (Updated
CC on 25-MAR-2003 to correct PN field.)
XX
XX Sequence 151 AA;
SQ
Query Match 88.1%; Score 683; DB 2; Length 151;
Best Local Similarity 90.1%; Pred. No. 2.2e-58;
Matches 136; Conservative 1; Mismatches 14; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQYGSANAALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQYGSANAALQ 60
QY 61 SPARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFRNATIDQWNAKNSDITVGYGG 120
DB 61 SPARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFRNATIDQWNAKNSDITVGYGG 120
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
RESULT 4
AAB36341
ID AAB36341 standard; protein; 151 AA.
XX
XX AAB36341;
AC
XX 26-FEB-2001 (first entry)
DT
XX
DE Salmonella enteritidis AgfA amino acid sequence SEQ ID NO:5.
XX
XX Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
XX vaccine; immune response; immunogen.
XX
XX Salmonella enteritidis.
XX
XX WO200060102-A2.
PN
XX 12-OCT-2000.
PD
XX
XX 05-APR-2000; 2000WO-CA000356.
PF
XX
XX 05-APR-1999; 99US-0127888P.
PR
XX
XX (UYVI-) UNIV VICTORIA.
PA
XX
XX White AP, Doran JL, Collinson SK, Kay WW;
PI
XX
XX WPI: 2000-572631/55.
DR
XX N-PSDB; AAC64617.
XX
XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 135; 139pp; English.
XX
XX The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:

```

```

CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
XX Sequence 151 AA;
SQ
Query Match 88.1%; Score 683; DB 3; Length 151;
Best Local Similarity 90.1%; Pred. No. 2.2e-58;
Matches 136; Conservative 1; Mismatches 14; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQYGSANAALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQYGSANAALQ 60
QY 61 SPARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFRNATIDQWNAKNSDITVGYGG 120
DB 61 SPARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFRNATIDQWNAKNSDITVGYGG 120
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
RESULT 5
AAW23570
ID AAW23570 standard; protein; 151 AA.
XX
XX AAW23570;
AC
XX 25-MAR-2003 (revised)
DT
XX 29-SEP-1997 (first entry)
DT
XX
DE Salmonella enteritidis 27655-3b agfA.
XX
XX Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody.
XX
XX Salmonella enteritidis.
OS
XX
XX Key Location/Qualifiers
FH Misc-difference 123
FT /note= "Encoded by GCC"
XX
XX US5635617-A.
PN
XX
XX 03-JUN-1997.
PD
XX
XX 26-APR-1994; 94US-00233788.
PF
XX
XX 26-APR-1993; 93US-00054452.
PR
XX
XX (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
PA
XX
XX Collinson SK, Kay WW, Doran JL;
PI

```

XX WPI; 1997-309886/28.
 XX N-PSDB; AAT74142.
 XX
 PT Isolated Salmonella gene agfA - used for diagnosis of Salmonella or
 PT enteropathogenic bacteria of the Enterobacteriaceae family.
 XX
 PS Example 2; Fig 7; 85pp; English.
 XX
 CC The present sequence represents agfA encoded by the full agfA gene
 CC derived from Salmonella enteritidis 27655-3b. The nucleic acid can be
 CC used to provide diagnostic assays for Salmonella and/or enteropathogenic
 CC bacteria of the family Enterobacteriaceae. It can also be used to provide
 CC proteins and antibodies which can be used for assays. The nucleic acid
 CC sequence can be used to provide probes or primers which can specifically
 CC hybridize to nucleic acid molecules from greater than 99% of Salmonella
 CC strains that are pathogenic to warm-blooded animals relative to nucleic
 CC acid molecules from virtually all other microbial organisms. (Updated on
 CC 25-MAR-2003 to correct PF field.)
 XX
 XX Sequence 151 AA;
 SQ
 Query Match 87.5%; Score 678; DB 2; Length 151;
 Best Local Similarity 89.4%; Pred No. 6.6e-58;
 Matches 135; Conservative 1; Mismatches 15; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
 QY 61 SDARKSETTITQSGYNGADYDQVLTVRVVTHEMAHARRNATIDQWNAKNSDITVGGYGG 120
 DB 61 SDARKSETTITQSGYNGADYDQVLTVRVVTHEMAHARRNATIDQWNAKNSDITVGGYGG 120
 QY 121 NNAALVNTASDSSVMVRQVGFNNATANQY 151
 DB 121 NNPAALVNTASDSSVMVRQVGFNNATANQY 151
 RESULT 6
 AAB36350
 ID AAB36350 standard; protein; 151 AA.
 AC AAB36350;
 XX
 XX 26-FEB-2001 (first entry)
 DE AgfA::PT3#5 amino acid sequence SEQ ID NO:20.
 XX Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
 KW vaccine; immune response; immunogen.
 XX
 XX Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 XX WO2000060102-A2.
 XX
 XX 12-OCT-2000.
 XX
 XX 05-APR-2000; 2000WO-CA000356.
 PF
 XX 05-APR-1999; 99US-0127888P.
 XX
 XX (UYVI-) UNIV VICTORIA.
 XX
 XX White AP, Doran JL, Collison SK, Kay WW;
 PI
 XX WPI; 2000-672631/65.
 DR N-PSDB; AAC64626.
 XX
 XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant agfA

PT protein useful for eliciting immune response in animal.
 XX Disclosure; Page 137; 139pp; English.
 PS
 XX
 CC The present invention describes a recombinant agfA gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/FAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant AgfA
 CC protein containing a replacement segment or segments of a foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;
 Query Match 85.4%; Score 662; DB 3; Length 151;
 Best Local Similarity 81.9%; Pred. No. 2.4e-56;
 Matches 136; Conservative 0; Mismatches 0; Indels 30; Gaps 2;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
 QY 61 SDARKSETTITQSGYNGADYDQVLTVRVVTHEMAHA-----FRNATIDQ 105
 DB 61 SDARK-----YDQVLTVRVVTHEMAHAGGADNSTIELTQNGFRNATIDQ 105
 QY 106 WNAKNSDITVGGYNGAALVNQTSDDSSVMVRQVGFNNATANQY 151
 DB 106 WNAKNSDITVGGYNGAALVNQTSDDSSVMVRQVGFNNATANQY 151
 RESULT 7
 AAB36353
 ID AAB36353 standard; protein; 151 AA.
 XX
 XX AAB36353;
 AC
 XX 26-FEB-2001 (first entry)
 DT
 XX AgfA::PT3#8 amino acid sequence SEQ ID NO:26.
 DE
 XX Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
 KW vaccine; immune response; immunogen.
 KW
 XX Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 OS
 XX WO2000060102-A2.
 XX
 XX 12-OCT-2000.
 XX
 XX 05-APR-2000; 2000WO-CA000356.
 PF
 XX

```

PR 05-APR-1999;      99US-0127888P.
XX
XX (UYVI-) UNIV VICTORIA.
XX
XX White AP, Doran JL, Collison SK, Kay WW;
XX
XX WPI; 2000-672631/65.
DR N-PSDB; AAC64629.
XX
XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 138; 139pp; English.
XX
XX The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SFF17/TRAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA, and AgfA-homologue fibrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fibrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fibrillar subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
XX Sequence 151 AA;
SQ
    Query Match          79.6%; Score 617; DB 3; Length 151;
    Best Local Similarity 73.6%; Pred. No. 5.7e-52;
    Matches 128; Conservative 0; Mismatches 0; Indels 46; Gaps 2
QY   1 MKLLKVAAPAAIIVSGSALAGVVPWGGGHHNGGNSGGPDSTLSIIYQVGSNAALAQ 60
Dd   1 MKLLKVAAPAAIIVSGSALAGVVPWGGGHHNGGNSGGPDSTLSIIYQVGSNAAL--- 57
QY   61 SDARKSETTITQSGYNGADYDQLVTRVVTVTHMAHA-----F 97
Dd   58 -----YDQLVTRVVTVTHMAHAGYNGADVQGQADNSTIELTQNGF 97
QY   98 RNNATIDQNAKNSDITVGQYGNNALVNQTASDSVMVVRQVFGNNATANQY 151
Dd   98 RNNATIDQNAKNSDITVGQYGNNALVNQTASDSVMVVRQVFGNNATANQY 151
RESULT 8
AAB36335
ID AAB36355 standard; protein; 151 AA.
XX
XX AAB36355;
XX
XX 26-FEB-2001 (first entry)
XX
XX AgfA::PT#10 amino acid sequence SEQ ID NO:30.
XX
XX Salmonella; agfA; chromosomal gene replacement; fimbrian; epitope;
KW vaccine; immune response; immunogen.

```

XX	Salmonella enteritidis.
OS	Escherichia coli.
OS	Synthetic.
XX	WO20060102-A2.
XX	12-OCT-2000.
PD	
PD	
XX	05-APR-2000; 2000WO-CA000356.
PF	
XX	05-APR-1999; 99US-012788BP.
PR	
XX	(UYVI-) UNIV VICTORIA.
PA	
PI	White AP, Doran JL, Collison SK, Kay WW;
XX	
DR	WPI; 2000-672631/65.
DR	N-PSDB; AAC64631.
XX	
PT	Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT	which encodes foreign epitope or antigen, expresses recombinant Agfa
PT	protein useful for eliciting immune response in animal.
XX	
PS	Disclosure, Page 139; 139pp; English.
XX	
CC	The present invention describes a recombinant agfa gene (I) where a
CC	segment of the gene has been replaced by a segment of a foreign DNA
CC	sequence which encodes a foreign epitope or antigen. Also described are:
CC	(1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC	assembly system of strains of Salmonella, Escherichia coli and
CC	Enterobacteriaceae for the production of fimbriae comprising recombinant
CC	Agfa, CsgA and Agfa-homologue fimbrin subunits, respectively; (2)
CC	directing recombination of a recombinant gene into the chromosome of the
CC	homologous species; (3) directing recombination of a recombinant gene
CC	back into the chromosome of the homologous species, replacing the native
CC	copy of that gene; and (4) eliciting an immune response in an animal,
CC	comprising separating an amino acid polymer comprising a recombinant Agfa
CC	protein-containing a replacement segment or segments of foreign amino
CC	acid sequence or sequences grown on a Salmonella, E. coli or
CC	Enterobacteriaceae host cell, from the host cell and introducing the
CC	polymer into the animal in conjunction with a carrier or diluent. (I) is
CC	useful for the expression of recombinant Agfa protein which is useful for
CC	eliciting an immune response in an animal. In a fimbrial presentation
CC	system the heterologous antigens are presented in high numbers (up to
CC	500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC	immunogenicity and adhesion properties relevant for an efficient live
CC	vaccine, the carrier fimbrial subunit proteins are usually strong
CC	immunogens, which may be important for directing an immune response
CC	against the inserted epitope, and hybrid fimbriae are easy and
CC	inexpensive to purify in large amount. The present sequence is given in
CC	the exemplification of the present invention
XX	
XX	Sequence 151 AA;
XX	
XX	Query Match 79.1%; Score 613; DB 3; Length 151;
XX	Best Local Similarity 74.6%; Pred. No. 1.4e-51;
XX	Matches 129; Conservative 0; Mismatches 0; Indels 44; Gaps 2
QY	i MKLLKVAFAAIIVSGSALAGVWPQGGGNGHNGGNSGDPSTLSIYQGSANAALAQ 60
Db	1 MKLLKVAFAAIIVSGSALAGVWPQGGGNGHNGGNSGDPSTLSIYQGSANAALAQ 60
QY	61 SDARKESETITQSGYGNGAD-----YDQLVTRVVTHMAFAFR 98
Db	61 SDARKESETITQSGYGNGADVQGCGADNSTIELIQGFRRNATYDQLVTRVVTHMAHA-- 118
QY	99 NNATIDQNNAKNSDITVGQYCGNNAAALVNQTASDSSVMVRQVFGFNNTANQY 151
Db	119 -----GNNAAALVNQTASDSSVMVRQVFGFNNTANQY 151

RESULT 9

CC the exemplification of the present invention
 XX Sequence 151 AA;
 SQ

Query Match 77.8%; Score 603; DB 3; Length 151;
 Best Local Similarity 80.8%; Pred. No. 1.3e-50;
 Matches 122; Conservative 4; Mismatches 25; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
 QY 61 SDARKSETTITQSGYNGADYDQLVTRVTHEMAHAFRNATIDQWNAKNSDITVGYGG 120
 DB 61 SDARKSETTITQSGYNGADYDQLVTRVTHEMAHAFRNATIDQWNAKNSDITVGYGG 120
 QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
 DB 121 LVRVTHEMAHAFRNATIDQWNAKNSDITVGYGG 120

RESULT 11
 AAB36349
 ID AAB36349 standard; protein; 151 AA.
 XX
 AC AAB36349;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE AgfA::PT3#4 amino acid sequence SEQ ID NO:18.
 XX
 KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 PN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay WW;
 XX
 DR WPI; 2000-672631/65.
 XX
 DR N-PSDB; AAC64625.
 XX

Recombinant agfA gene having a segment replaced by a foreign DNA sequence which encodes foreign epitope or antigen, expresses recombinant AgfA protein useful for eliciting immune response in animal.
 PS Disclosure; Page 136; 139pp; English.
 XX

The present invention describes a recombinant agfA gene (I) where a segment of the gene has been replaced by a segment of a foreign DNA sequence which encodes a foreign epitope or antigen. Also described are: (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended assembly system of strains of Salmonella, Escherichia coli and Enterobacteriaceae for the production of fimbriae comprising recombinant AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2) directing recombination of a recombinant gene into the chromosome of the homologous species; (3) directing recombination of a recombinant gene back into the chromosome of the homologous species, replacing the native copy of that gene; and (4) eliciting an immune response in an animal comprising separating an amino acid polymer comprising a recombinant AgfA protein containing a replacement segment or segments of foreign amino acid sequence or sequences grown on a Salmonella, E. coli or

CC Enterobacteriaceae host cell, from the host cell and introducing the polymer into the animal in conjunction with a carrier or diluent. (I) is useful for the expression of recombinant AgfA protein which is useful for eliciting an immune response in an animal. In a fimbrial presentation system the heterologous antigens are presented in high numbers (up to 500,000 copies/cell), the hybrid fimbrin protein possesses both the immunogenicity and adhesion properties relevant for an efficient live vaccine, the carrier fimbrial subunit proteins are usually strong immunogens, which may be important for directing an immune response against the inserted epitope, and hybrid fimbriae are easy and inexpensive to purify in large amount. The present sequence is given in the exemplification of the present invention
 XX Sequence 151 AA;
 SQ

Query Match 77.5%; Score 601; DB 3; Length 151;
 Best Local Similarity 80.8%; Pred. No. 2e-50; Mismatches 25; Indels 0; Gaps 0;
 Matches 122; Conservative 4;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
 QY 61 SDARKSETTITQSGYNGADYDQLVTRVTHEMAHAFRNATIDQWNAKNSDITVGYGG 120
 DB 61 SDARKSETTITQSGYNGADYDQLVTRVTHEMAHAFRNATIDQWNAKNSDITVGYGG 120
 QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
 DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 12
 AAB36352
 ID AAB36352 standard; protein; 151 AA.
 XX
 AC AAB36352;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE AgfA::PT3#7 amino acid sequence SEQ ID NO:24.
 XX
 KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 PN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay WW;
 XX
 DR WPI; 2000-672631/65.
 XX
 DR N-PSDB; AAC64628.
 XX

Recombinant agfA gene having a segment replaced by a foreign DNA sequence which encodes foreign epitope or antigen, expresses recombinant AgfA protein useful for eliciting immune response in animal.
 PS Disclosure; Page 138; 139pp; English.
 XX

The present invention describes a recombinant agfA gene (I) where a segment of the gene has been replaced by a segment of a foreign DNA sequence which encodes a foreign epitope or antigen. Also described are: (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended assembly system of strains of Salmonella, Escherichia coli and Enterobacteriaceae for the production of fimbriae comprising recombinant AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2) directing recombination of a recombinant gene into the chromosome of the homologous species; (3) directing recombination of a recombinant gene back into the chromosome of the homologous species, replacing the native copy of that gene; and (4) eliciting an immune response in an animal comprising separating an amino acid polymer comprising a recombinant AgfA protein containing a replacement segment or segments of foreign amino acid sequence or sequences grown on a Salmonella, E. coli or

CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant AgfA
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC the exemplification of the present invention

XX Sequence 151 AA;

Query Match 77.4%; Score 600; DB 3; Length 151;
 Best Local Similarity 81.5%; Pred. No. 2.5e-50;
 Matches 123; Conservative 3; Mismatches 25; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQGSANAALALQ 60
 Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQGSANAALALQ 60
 QY 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAFAFRNATIDQWNAKNSDITVGYGG 120
 Db 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAFAFRNATIDQWNAKNSDITVTRVVT 120
 QY 121 NNAALVNQTASDSSVWVRQVGFNNATANQY 151
 Db 121 HEMAHANQTASDSSVWVRQVGFNNATANQY 151

RESULT 13

AAAB36348
 ID AAB36348 standard; protein; 151 AA.

XX AAB36348;

XX 26-FEB-2001 (first entry)

DE AgfA::PT3#3 amino acid sequence SEQ ID NO:16.

KW Salmonella; agfA; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.

XX Salmonella enteritidis.

OS Escherichia coli.

XX Synthetic.

XX WO200060102-A2.

XX 12-OCT-2000.

XX 05-APR-2000; 2000WO-CA000356.

XX 05-APR-1999; 99US-0127889P.

XX (UYVI-) UNIV VICTORIA.

XX White AP, Doran JL, Collison SK, Kay WW;

XX WPI; 2000-672631/65.

DR N-PSDB; AAC64624.
 XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant AgfA
 PT protein useful for eliciting immune response in animal.
 XX Disclosure; Page 136; 139pp; English.

CC The present invention describes a recombinant agfA gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant AgfA
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

XX Sequence 151 AA;

Query Match 73.3%; Score 568; DB 3; Length 151;
 Best Local Similarity 80.1%; Pred. No. 3.3e-47;
 Matches 121; Conservative 4; Mismatches 26; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQGSANAALALQ 60
 Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAFAFRNATIDQWNAKNSDITVGYGG 120
 Db 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAFAFRNATIDQWNAKNSDITVGYGG 120

QY 121 NNAALVNQTASDSSVWVRQVGFNNATANQY 151

Db 121 NNAALVNQTASDSSVWVRQVGFNNATANQY 151

RESULT 14

AAAB36343
 ID AAB36343 standard; protein; 151 AA.

XX AAB36343;

XX 26-FEB-2001 (first entry)

XX Escherichia coli CsgA amino acid sequence SEQ ID NO:7.

XX Salmonella; agfA; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.

OS Escherichia coli.

XX WO200060102-A2.

XX 12-OCT-2000.

```
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
FI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
XX N-PSDB; AAC64619.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 135; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 151 AA;
Query Match 65.7%; Score 509; DB 3; Length 151;
Best Local Similarity 68.2%; Pred. No. 1.8e-41;
Matches 103; Conservative 15; Mismatches 33; Indels 0; Gaps 0;
QY 1 MKLLKVAFAIAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALAQ 60
Db 1 MKLLKVAIAIAIVFSGSALAGVVPQYGGGNGHGGNSGPNSELNTIYQYGGNSALAQ 60
QY 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFRNNTATIDWNKNSDITVGYGG 120
Db 61 TDARNSDLTITQHGNGGADVGQGSDDSIDLTQRFNGNSATLDWNGKNSMTVKQFGG 120
QY 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151
Db 121 GNGAAVDQTASNSNVTVQVGFNNATAHQY 151
RESULT 15
ABR82651
XX ABR82651 standard; protein; 151 AA.
XX
AC ABR82651;
XX
DT 04-DEC-2003 (first entry)
XX
DE E. coli CsgA subunit 15 kDa protein.
XX
```

Plasma protein; immune response; antibacterial; vaccine; gene therapy.

Escherichia coli.

WO2003064446-A2.

07-AUG-2003.

30-JAN-2003; 2003WO-EP000943.

31-JAN-2002; 2002GB-00002275.

(HANS-) HANSA MEDICAL RES AB.

Bjoerck L, Olsen A, Wikstroem M, Herwald H;

WPI; 2003-646136/61.

N-PSDB; ACF36153.

New isolated peptide capable of binding a mammalian plasma protein, useful in the manufacture of a medicament for the prevention and/or treatment of a bacterial infection, such as Escherichia coli, Salmonella or Shigella infections.

Disclosure; Page 41-42; 42pp; English.

The invention relates to an isolated peptide capable of binding a mammalian plasma protein or of generating an immune response in a mammal selected from sequences shown in ABR82642, ABR82648-49. The peptide or antibody is useful for treating a bacterial infection in a human or animal or in the manufacture of a medicament for the prophylactic treatment of a bacterial infection, such as Escherichia coli, Salmonella or Shigella infection. The peptide that is immobilized on a solid support is also useful as a reagent for determining the ability of a plasma protein to bind to bacteria. The present sequence represents an E. coli 15 kDa protein

Sequence 151 AA;

Query Match 65.0%; Score 504; DB 7; Length 151;

Best Local Similarity 67.5%; Pred. No. 5.5e-41;

Matches 102; Conservative 15; Mismatches 34; Indels 0; Gaps 0;

QY 1 MKLLKVAFAIAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALAQ 60

Db 1 MKLLKVEIAIAIVFSGSALAGVVPQYGGGNGHGGNSGPNSELNTIYQYGGNSALAQ 60

QY 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFRNNTATIDWNKNSDITVGYGG 120

Db 61 TDARNSDLTITQHGNGGADVGQGSDDSIDLTQRFNGNSATLDWNGKNSMTVKQFGG 120

QY 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151

Db 121 GNGAAVDQTASNSNVTVQVGFNNATAHQY 151

Search completed: August 2, 2004, 14:48:29

Job time : 45.9 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:40:48 ; Search time 12 Seconds
(without alignments)
649.627 Million cell updates/sec

Title: US-09-543-407-28
Perfect score: 775
Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMYRQVGFNNATANYQ 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

- 1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/2/iaa/6C_COMB.pep.*
- 6: /cgn2_6/ptodata/2/iaa/6D_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	678	87.5	151	1	US-08-233-788A-59
2	498	64.3	120	1	US-08-233-788A-57
3	85	11.0	975	4	US-08-328-352-4764
4	84.5	10.9	738	3	US-08-864-038A-3
5	84	10.8	943	4	US-09-056-556-204
6	84	10.8	943	4	US-09-072-596-199
7	84	10.8	943	4	US-09-477-135A-131
8	84	10.8	943	4	US-09-072-967-204
9	82	10.6	2123	3	US-08-968-685A-10
10	81.5	10.5	558	4	US-08-252-991A-30983
11	80	10.3	339	4	US-08-252-991A-32096
12	80	10.3	1207	4	US-08-489-039A-11518
13	79.5	10.3	745	4	US-08-336-115C-6
14	79.5	10.3	1415	4	US-08-252-991A-26438
15	78.5	10.1	714	4	US-08-841-786-4
16	78.5	10.1	3241	4	US-08-841-786-1
17	78	10.1	892	4	US-08-336-447A-5
18	78	10.1	918	4	US-08-200-850E-1
19	77.5	10.0	415	4	US-08-025-769B-280
20	77.5	10.0	518	3	US-08-043-123-2
21	77	9.9	1160	3	US-08-808-599A-24
22	76.5	9.9	873	4	US-08-336-447A-13
23	75.5	9.7	211	1	US-08-276-852-34
24	75.5	9.7	211	1	US-08-133-011-16
25	75.5	9.7	211	1	US-08-322-730A-16
26	75.5	9.7	211	1	US-08-387-874-16
27	75.5	9.7	211	1	US-08-899-575-34

28 75:5 9:7 211 1 US-08-899-575-34 Sequence 34, Appl
29 75:5 9:7 211 2 US-08-383-619-16 Sequence 16, Appl
30 75:5 9:7 211 3 US-08-907-739-16 Sequence 16, Appl
31 75:5 9:7 211 4 US-09-729-597-16 Sequence 16, Appl
32 75:5 9:7 211 5 PCT-US93-08364-16 Sequence 16, Appl
33 75:5 9:7 211 5 PCT-US95-08743-34 Sequence 34, Appl
34 75:5 9:7 238 4 US-09-495-880A-42 Sequence 42, Appl
35 75:5 9:7 244 2 US-08-553-497A-22 Sequence 22, Appl
36 75:5 9:7 266 4 US-09-495-880A-26 Sequence 26, Appl
37 75:5 9:7 293 3 US-08-438-745-6 Sequence 4, Appl
38 75:5 9:7 293 3 US-08-438-745-6 Sequence 6, Appl
39 75:5 9:7 293 3 US-09-219-019-4 Sequence 4, Appl
40 75:5 9:7 293 3 US-09-219-019-6 Sequence 6, Appl
41 75:5 9:7 293 5 PCT-US94-05669A-4 Sequence 4, Appl
42 75:5 9:7 293 5 PCT-US94-05669A-6 Sequence 6, Appl
43 75:5 9:7 432 4 US-09-403-089A-1 Sequence 1, Appl
44 75:5 9:7 461 2 US-08-463-587A-26 Sequence 26, Appl
45 75:5 9:7 461 2 US-08-463-667A-4 Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-08-233-788A-59
; Sequence 59, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, Karen S.
; APPLICANT: Clouthier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; TITLE OF INVENTION: OF SALMONELLA
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,788A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 59:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 151 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-233-788A-59

Query Match 87.5%; Score 678; DB 1; Length 151;
Best Local Similarity 89.4%; Pred No. 4e-61;
Matches 135; Conservative 1; Mismatches 15; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVYVQWGGGNGHNGSGPDSTLSIYQYGSANALALQ 60
|||||

Db 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQVGSAAALALQ 60
QY 61 SDARKSETTITQSGYGGAGDYDQLVTRVTVTHEMAHAFRNATIDQWNAKNSDITVGYGG 120
Db 61 SDARKSETTITQSGYGGAGDYDQADNSTIELTQNGFRNATIDQWNAKNSDITVGYGG 120
QY 121 NNAALVNTASDSSVMVQVGFQGNATANQY 151
Db 121 NNPALVNTASDSSVMVQVGFQGNATANQY 151

RESULT 2

US-08-233-788A-57

; Sequence 57, Application US/08233788A

; Patent No. 5635617

; GENERAL INFORMATION:

; APPLICANT: Doran, James L.

; APPLICANT: Kay, William W.

; APPLICANT: Collinson, Karen S.

; APPLICANT: Clouchier, Sharon C.

; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION

; TITLE OF INVENTION: OF SALMONELLA

; NUMBER OF SEQUENCES: 61

; CORRESPONDENCE ADDRESS:

; STREET: 6300 Columbia Center, 701 Fifth Avenue

; CITY: Seattle

; STATE: Washington

; COUNTRY: U.S.A.

; ZIP: 98104-7092

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/233,788A

; FILING DATE: 26-APR-1994

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: King, Joshua

; REGISTRATION NUMBER: 35,570

; REFERENCE/DOCKET NUMBER: 920043.403C2

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (206) 622-4900

; TELEFAX: (206) 682-6031

; TELEX: 3723836 SEEDANBERRY

; INFORMATION FOR SEQ ID NO: 57:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 120 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-233-788A-57

Query Match

Best Local Similarity 64.3%; Score 498; DB 1; Length 120;

Matches 97; Conservative 1; Mismatches 14; Indels 0; Gaps 0;

QY 22 VVPQWGGGNGHNGGNSGPDSTLSIYQVGSAAALALQSDARKSETTITQSGYGGAGDY 81

Db 1 VVPQWGGGNGHNGGNSGPDSTLSIYQVGSAAALALQSDARKSETTITQSGYGGAGDY 60

QY 82 DQLVTRVTHEMAHAFRNATIDQWNAKNSDITVQVGGNNAALVNTASDS 133

Db 61 QGQADNSTIELTQNGFRNATIDQWNAKNSDITVQVGGNNAALVNTASDS 112

RESULT 3

US-09-328-352-4764

; Sequence 4764, Application US/09328352

; Patent No. 6562958

; GENERAL INFORMATION:

; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03FA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4764
; LENGTH: 975
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-4764

Query Match 11.0%; Score 85; DB 4; Length 975;

Best Local Similarity 25.3%; Pred No. 4.2;

Matches 38; Conservative 18; Mismatches 52; Indels 42; Gaps 7;

QY 15 SGSALAGVVPQWGGGNGHNGG-GNSSGPDSTLSIYQYG-----SANA 55

Db 300 AGNIA-----SGNGEHNIGGNGGDDVDITAPITGVNLISGNSFTLIGNSSSSSVNT 353

QY 56 ALALQSDARKSETTI-----TQSGYGGAGDYDQLVTRVTHEMAHAFRNATIDQWNA 108

Db 354 APTTSTNTVNDNTIDNGSGGTGSGGNGSG-DGLLNGAASNGEH-----NYGIGNGNG 408

QY 109 KNSDIT-----VQYCGNNAALVNTASDS 133

Db 409 DQVDITSPITGIFNFGNSFSLIGNSSSSS 438

RESULT 4

US-08-864-038A-3

; Sequence 3, Application US/08864038A

; Patent No. 6001592

; GENERAL INFORMATION:

; APPLICANT: Kunio NAKASHIMA et al.

; TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE CDNA, VECTOR

; TITLE OF INVENTION: CONTAINING SAID CDNA, HOST CELLS TRANSFORMED WITH SAID

; TITLE OF INVENTION: VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING

; TITLE OF INVENTION: SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBODY

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSER: 812-5 Hirano

; STREET: Isehinden

; CITY: Tsu-city

; STATE: Mie-prefecture

; COUNTRY: JAPAN

; ZIP: 514-01

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: Microsoft Windows 95

; SOFTWARE: Word Perfect 6.1

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/864,038A

; FILING DATE: May 28, 1997

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: JP 8-184459

; FILING DATE: 15-July-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: C. Bruce Hamburg

; REGISTRATION NUMBER: 22,389

; REFERENCE/DOCKET NUMBER: F-5610

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212)986-2340

; TELEFAX: (212)953-7733

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 738

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

[illegible]

PROTEIN-106 POLYPEPTIDE GENE SEQUENCE AND USES THEREOF

US-08-968-685A-10

Query Match
Best Local Similarity 10.6%; Score 82; DB 3; Length 2123;
Matches 31; Conservative 17; Mismatches 38; Indels 34; Gaps 5;

QY 1 MKLLKVAFAAIV-----SGSALAGVVPQWGGGNGGNSGPDSTLSIYQYGSANAA 56
DB 44 LSFARIALAVLIVGATINGSAYAGIGISEADGG--KGGANARG-DKSIAT--GDIQA 97
QY 57 LALQSDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFRNNATIDQWNAKNSDITVG 116
DB 98 LGSQSI-----IGNKIVHNSNNANIGAKASGNSIAIG 133

RESULT 10
US-09-252-991A-30983
; Sequence 30983, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30983
; LENGTH: 556
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30983

Query Match
Best Local Similarity 10.5%; Score 81.5; DB 4; Length 558;
Matches 27; Conservative 11; Mismatches 39; Indels 11; Gaps 2;

QY 16 GSALAGV-----VPQWGGGNGGNSGPDSTLSIYQYGSANAAALQSDARKSETTI 70
DB 220 GAFAVGEGALVQQLGGHALDGGADAG-----AVHEGHAFAQALVRSDEEAGGAVE 273
QY 71 TQSGYNGADYDQLVTRVVTHEMAHAFR 98
DB 274 VHHAGGRGLDPHLLQRAATHAVARAER 301

RESULT 11
US-09-252-991A-32096
; Sequence 32096, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 32096
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-32096

Query Match
10.3%; Score 80; DB 4; Length 339;

Best Local Similarity 20.1%; Pred. No. 3.3;
Matches 31; Conservative 34; Mismatches 61; Indels 28; Gaps 6;

QY 5 KYAFAAIVVSGSALAGVVPQWGG-----GNGHNGGNSGPDSTLSIYQYGSANAA 57
DB 58 KVSNFOTL--NNASVSGSIKDGNGVNVVAAGDNNQQAALASADASFVFGTATAS- 114
QY 58 ALQSDARKSETTITQSGYNGA-DYDQLVTRVVTHEMAHAFRNNATIDQWNAKNSDI 113
DB 115 -----TSVLQSGYGTNNYSNPNTASLSNSANNVSGNLGVNVAAGNFNQKNDL 164
QY 114 TV-----GQYGGNNAALVNOTASDSSVMVQVFG 143
DB 165 AAASVNGQYSTAGSAASTGTGTNTVNSANYAIG 198

RESULT 12
US-09-489-039A-11518
; Sequence 11518, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 11518
; LENGTH: 1207
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-11518

Query Match
10.3%; Score 80; DB 4; Length 1207;
Best Local Similarity 23.5%; Pred. No. 18;
Matches 36; Conservative 23; Mismatches 66; Indels 28; Gaps 5;

QY 21 GVPQWGGGNNH--GGGNSGPDSTLSIYQYGS-----ANAAALQSDARKSETTITOS 73
DB 482 GPVMSQPAYNNHQFSGAINGAESMISTFSLGSLASGVSSQAALQSQKSEAFQSTLGR 541
QY 74 GYNGADYDQLVTRV-----VTHEMAHAFRNNATIDQWNAKNSDITVQ 117
DB 542 GFGDGVSDQAYSRLSNIGRNVSSQNTAQSQLNQAKNFMDFQVDD----SHSDAVKGA 598
QY 118 YG--GNNAAALVNOTASDSSVMVQVFGNNATA 148
DB 599 FAVQATGTLDDVQAAASMLMPVVGKARAAKAAA 631

RESULT 13
US-09-336-115C-6
; Sequence 6, Application US/09336115C
; Patent No. 6576244
; GENERAL INFORMATION:
; APPLICANT: Weltzin, Richard A.
; APPLICANT: Guy, Bruno
; TITLE OF INVENTION: Lt and Ct in Parenteral Immunization
; TITLE OF INVENTION: Methods Against Helicobacter Infection
; FILE REFERENCE: 06132/055002
; CURRENT APPLICATION NUMBER: US/09/336,115C
; CURRENT FILING DATE: 1999-06-18
; PRIOR APPLICATION NUMBER: US 09/100,258
; PRIOR FILING DATE: 1998-06-19
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 745
; TYPE: PRT
; ORGANISM: Helicobacter pylori

```

; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)....(20)
;
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 721
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-336-115C-6

Query Match          10.3%; Score 79.5; DB 4; Length 745;
Best Local Similarity 20.5%; Pred. No. 11;
Matches 35; Conservative 27; Mismatches 70; Indels 39; Gaps 6;

QY 8 AFPAIWSGSALAGVVPQW-----GGGNHNGGN----- 37
DQ 102 AYQAVFLAINAAVGL--WNTIGYAVCMGNGTSGPGSVIFNDQPGQDSTQITCNRFE 158
QY 38 SSGPDSSTLSIYQYGSANAALALQSDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAF 97
DQ 159 STPGKSMISIDBFKXLEAYQIIQALKXQSGFPPLG-GNCTK----VSVNRYNECRQTA 213
QY 98 RNNATIDOW-NAKNSDITVQYGGNNAALVNCTASDSSVMVRQVGFGNAT 147
DQ 214 DINGGVYQFCKAKNGSSSSNGSGSTQTATTITQDGVTTITTYNNKAT 264

RESULT 14
US-09-252-991A-26438
; Sequence 26438, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 26438
; LENGTH: 1415
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-26438

Query Match          10.3%; Score 79.5; DB 4; Length 1415;
Best Local Similarity 28.9%; Pred. No. 25;
Matches 44; Conservative 17; Mismatches 74; Indels 17; Gaps 9;

QY 10 AAIIVSGSALAGVVPQWGGG--NHNGG--NSSGPDSTLSIYQYGSANAALALQSDA-- 63
DQ 930 ADFAISGQ-LKDHASHYGAGGLVGNRGLIRSSGSGTSLSGHGMNLGLVGYSSAGG 988
QY 64 -RKSETTITQSGYNGADYDQLVTRVVTHEMAHAFRNNATIDQWNAKNSDITVQYGN- 121
DQ 989 LADVASVDVSGNGQGLYGLGLGNVNSGTAHATASGKV----RGTDAAELGGLIGNL 1044
QY 122 NAALVNQTA-SDSSVMV-RQVG--FGNNATAN 149
DQ 1045 NAAINNASHGDVSLQAGRYLGLLGHCHNQAGN 1076

RESULT 15
US-09-841-786-4
; Sequence 4, Application US/09841786
; Patent No. 6669940
; GENERAL INFORMATION:
; APPLICANT: NAGARAJA, T. G.
; APPLICANT: STEWART, GEORGE C.
; APPLICANT: NARAYANAN, SANJEEV K.

```

```

; APPLICANT: CHENGAPPA, M. M.
; TITLE OF INVENTION: RECOMBINANT FUSOBACTERIUM NECROPHORUM LEUKOTOXIN
; TITLE OF INVENTION: VACCINE AND PREPARATION THEREOF
; FILE REFERENCE: 30296
; CURRENT APPLICATION NUMBER: US/09/841,786
; CURRENT FILING DATE: 2001-04-24
; PRIOR APPLICATION NUMBER: 09/558,257
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 714
; TYPE: PRT
; ORGANISM: Fusobacterium necrophorum
US-09-841-786-4

Query Match          10.1%; Score 78.5; DB 4; Length 714;
Best Local Similarity 22.8%; Pred. No. 13;
Matches 31; Conservative 16; Mismatches 56; Indels 33; Gaps 6;

QY 12 IVVSGSALAGVVPQWGGG-----NHNGGN-----SSGPDSTLSIYQYGSANAALALQSDAR 64
DQ 35 VTSSDSTFVGA---WGGSAALQWNHIGSGNSNISAGLAGAAAANNIQTSLVVKNSDIR 91
QY 65 KSE-----TTITQSGYNGADY---DQLVTRVVTHEMAHAFRNNAT 102
DQ 92 NANKFKNNALSGTQVAAGAGLEAVKESG-GQCKSYLLCTASINLVNNEVSAKSENNTV 150
QY 103 IDQWNAKNSDITVQY 118
DQ 151 AGESESQKMDVDVTAY 166

Search completed: August 2, 2004, 14:58:36
Job time : 12 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:54:48 ; Search time 36.8 Seconds
(without alignments)

1287.123 Million cell updates/sec

Title: US-09-543-407-28

Perfect score: 775

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVFGNNATANCY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query No.	Score	Match	Length	DB ID	Description
1	506	65.3	151	12	US-09-741-873B-4	Sequence 4, Appli
2	506	65.3	151	12	US-09-741-873B-4	Sequence 4, Appli
3	428	55.2	131	12	US-09-741-873B-2	Sequence 2, Appli
4	428	55.2	131	12	US-09-741-873B-2	Sequence 2, Appli
5	95	12.3	400	16	US-10-437-963-186417	Sequence 186417,
6	92	11.9	1074	12	US-10-282-122A-50616	Sequence 50616, A
7	89	11.5	354	10	US-09-820-843A-21	Sequence 21, Appli
8	87.5	11.3	249	9	US-09-730-374-3	Sequence 3, Appli
9	87.5	11.3	249	16	US-10-704-206-3	Sequence 3, Appli
10	87.5	11.3	445	15	US-10-369-493-20638	Sequence 20638, A
11	87.5	11.3	1448	9	US-10-408-765A-998	Sequence 998, App
12	87	11.2	597	6	US-09-793-306-146	Sequence 146, App
13	86	11.1	974	12	US-10-282-122A-49999	Sequence 49999, A
14	85.5	11.0	2204	10	US-10-282-122A-64364	Sequence 64364, A
15	85	11.0	254	10	US-09-880-748-905	Sequence 905, App

16 85 11.0 254 12 US-10-293-418-905 Sequence 905, App
17 84.5 10.9 928 12 US-10-282-122A-54647 Sequence 54647, A
18 84.5 10.9 978 12 US-10-282-122A-54379 Sequence 54379, A
19 84 10.8 943 9 US-09-995-634-131 Sequence 131, App
20 84 10.8 943 10 US-09-997-182-131 Sequence 131, App
21 84 10.8 943 10 US-09-997-181-131 Sequence 131, App
22 84 10.8 943 14 US-10-193-002-199 Sequence 199, App
23 84 10.8 943 14 US-10-084-843-204 Sequence 204, App
24 84 10.8 3300 12 US-10-282-122A-64369 Sequence 64369, A
25 83.5 10.8 154 16 US-10-437-963-162284 Sequence 162284,
26 83.5 10.8 288 12 US-10-434-599-283675 Sequence 283675,
27 83 10.7 613 12 US-10-282-122A-72467 Sequence 72467, A
28 83 10.7 793 12 US-10-282-122A-50037 Sequence 50037, A
29 82.5 10.6 145 16 US-10-437-963-147748 Sequence 147748,
30 82.5 10.6 319 12 US-10-244-596-32 Sequence 32, Appli
31 82.5 10.6 319 12 US-10-244-596-33 Sequence 33, Appli
32 82.5 10.6 580 12 US-10-647-057-4 Sequence 4, Appli
33 82 10.6 186 12 US-10-282-122A-49412 Sequence 49412, A
34 82 10.6 369 12 US-10-425-114-56041 Sequence 56041, A
35 82 10.6 486 12 US-10-424-599-275468 Sequence 275468,
36 82 10.6 507 12 US-10-425-114-57763 Sequence 57763, A
37 82 10.6 2122 9 US-09-813-214A-9 Sequence 9, Appli
38 81.5 10.5 485 12 US-10-282-122A-54094 Sequence 54094, A
39 81 10.5 224 9 US-09-738-626-4894 Sequence 4894, Ap
40 81 10.5 341 12 US-10-424-599-246651 Sequence 246651,
41 81 10.5 703 16 US-10-437-963-108981 Sequence 108981,
42 81 10.5 812 15 US-10-369-493-9134 Sequence 9134, Ap
43 80.5 10.4 342 16 US-10-437-963-187064 Sequence 187064,
44 80.5 10.4 628 12 US-10-282-122A-53269 Sequence 53269, A
45 80 10.3 246 10 US-09-880-748-1847 Sequence 1847, Ap

ALIGNMENTS

RESULT 1
US-09-741-873B-4
; Sequence 4, Application US/09741873B
; Publication No. US20020081722A1
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; CURRENT FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-4

Query Match 65.3%; Score 506; DB 12; Length 151;
Best Local Similarity 67.5%; Pred. No. 2.3e-42;
Matches 102; Conservative 16; Mismatches 33; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNSGSPDSTLSIYQVGSANAALAQ 60

```
Db 1 MKLLKVAATAIVFSGSAGVVPQYGGGNGHGGGNSGPNSELNIYQYGGNSALALQ 60
QY 61 SDARKSETTITQSGYGCAGDYDVLVTRVVTHEMAHAFFNNATIDOWNAKNSDITVGOYGG 120
Db 61 TDARNSDITITQHGNGGAGVGGSDSSIDLTRGFGNSALDQWNGKNSMTVKQFGG 120
QY 121 NNAALVNQASDSSVMVRQVGFNNATANQY 151
Db 121 NGGAADVDTASNSVNVTVQVFGNNATAHQY 151

RESULT 2
US-09-741-873B-4
; Sequence 4, Application US/09741873B
; Publication No. US20040096965A9
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; CURRENT FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1991-11-03
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-4
Query Match 65.3%; Score 506; DB 12; Length 151;
Best Local Similarity 67.5%; Pred. No. 2.3e-42;
Matches 102; Conservative 16; Mismatches 33; Indels 0; Gaps 0;

QY 1 MKLLKVAATAIVVSGSALAGVVPQWGGGNGHGGGNSGPDSTLSIYQYGSANAALALQ 60
Db 1 MKLLKVAATAIVFSGSAGVVPQYGGGNGHGGGNSGPNSELNIYQYGGNSALALQ 60
QY 61 SDARKSETTITQSGYGCAGDYDVLVTRVVTHEMAHAFFNNATIDOWNAKNSDITVGOYGG 120
Db 61 TDARNSDITITQHGNGGAGVGGSDSSIDLTRGFGNSALDQWNGKNSMTVKQFGG 120
QY 121 NNAALVNQASDSSVMVRQVGFNNATANQY 151
Db 121 NGGAADVDTASNSVNVTVQVFGNNATAHQY 151

RESULT 3
US-09-741-873B-2
; Sequence 2, Application US/09741873B
; Publication No. US20020081722A1
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; CURRENT FILING DATE: 2003-04-04
```

```
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1998-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-2
Query Match 55.2%; Score 428; DB 12; Length 131;
Best Local Similarity 64.1%; Pred. No. 1.1e-34;
Matches 84; Conservative 15; Mismatches 32; Indels 0; Gaps 0;

QY 21 GVVPQWGGGNGHGGGNSGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYGCAG 80
Db 1 GVVPQYGGGNGHGGGNSGPNSELNIYQYGGNSALALQTDARNSDITITQHGNGGAG 60
QY 81 YDQLVTRVVTHEMAHAFFNNATIDOWNAKNSDITVGOYGSANAALVNQASDSSVMVRQV 140
Db 61 VQGSDDSSIDLTRGFGNSALDQWNGKNSMTVKQFGGNGAADVDTASNSVNVTVQV 120
QY 141 GFGNNATANQY 151
Db 121 GFGNNATAHQY 131

RESULT 4
US-09-741-873B-2
; Sequence 2, Application US/09741873B
; Publication No. US20040096965A9
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; CURRENT FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-2
Query Match 55.2%; Score 428; DB 12; Length 131;
```

Best Local Similarity 64.1%; Pred. No. 1.1e-34; Matches 84; Conservative 15; Mismatches 32; Indels 0; Gaps 0;

QY 21 GVVPMQGGGNNHGGNSGPDSTLSIYQYGSANAALQSDARKSETTITQSGYGNAGD 80
Db 1 GVVPMQGGGNNHGGNSGPNSELATYQYGGNSALALQTDARNSDLIITQHGGNGAD 60
QY 81 YDQIVRVVTHEMAHAFRNATIDQNAKNSDITVQYGGNNAALVNTASDSSVMVROV 140
Db 61 VGGGSDSSIDLQRFQFGNSATLDQNGKNSEMTVKQFGGNGAAVDQTSNNSVNVTVQ 120
QY 141 GFGNNATANOY 151
Db 121 GFGNNATAHOY 131

RESULT 5

US-10-437-963-186417
; Sequence 186417, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 186417
; LENGTH: 400
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_83218C.1.pap
US-10-437-963-186417

Query Match 12.3%; Score 95; DB 16; Length 400;
Best Local Similarity 23.6%; Pred. No. 0.52;
Matches 37; Conservative 20; Mismatches 62; Indels 38; Gaps 5;

QY 14 VSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYQYGSANAALQSDARKSETTITQ 73
Db 252 MNGSSSSSMFQLGGGGMSNGNS-----FSASMAAFMDEEAAMWATAVAPP 300
QY 74 GYNGADYDQLV-----TRVVTHEMAHAFRNATIDQNAKNSDIT 114
Db 301 GMGLAGNDVQAQQQQQVLQDAVGVAPTLLXVHGGGAATAGAMVVDX-SVMVVDVS 359
QY 115 VQYGGNNAALVNTASDSSVMVROVGF--GNNTAN 149
Db 360 SAVYGGATA-----TAFDLDMVESCGMFCGGGAGN 391

RESULT 6

US-10-282-122A-50616
; Sequence 50616, Application US/10282122A
; Publication No. US20040029129A1

; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John

; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 50616
; LENGTH: 1074
; TYPE: PRT
; ORGANISM: Burkholderia mallei
US-10-282-122A-50616

Query Match 11.9%; Score 92; DB 12; Length 1074;
Best Local Similarity 28.1%; Pred. No. 3.5;
Matches 39; Conservative 21; Mismatches 69; Indels 10; Gaps 5;

QY 13 VVGSALAGVVPQWGGGNNHGGNSGPDSTLSIYQYGSANAALQSDARKSETTITQ 72
Db 22 IASGULLA--APAWADTVSPGTDNVYGVDAIDPGVSTNQGTAYGAQAGAKVTGSYNTA 79
QY 73 SCYNGADYDQLVTRVVTHEMAHAFRNATIDQNAKNS---DITVQYGGNNAALVNOT 129
Db 80 IGYQAGQNVN-VIDIVSIGKQATASANDAIAIG-TNKASGPADIYMGILNAGAG--STT 136
QY 130 ASDSSVM--VRQVGFQNA 146
Db 137 SPDGTVTLGIRNMGLESA 155

RESULT 7

US-09-820-843A-21
; Sequence 21, Application US/09820843A
; Publication No. US20030039963A1

; GENERAL INFORMATION:

; APPLICANT: Council of Scientific and Industrial Research

; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTE

; FILE REFERENCE: Q63915

; CURRENT APPLICATION NUMBER: US/09/820,843A

; CURRENT FILING DATE: 2001-03-30

; NUMBER OF SEQ ID NOS: 118

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 21

; LENGTH: 354

; TYPE: PRT

; ORGANISM: M. tuberculosis

; FEATURE:

; NAME/KEY: misc_feature

```
; OTHER INFORMATION: PPE
; NAME/KEY: misc.feature
; OTHER INFORMATION: GI|1781260
US-09-820-843A-21

Query Match      11.5%; Score 89; DB 10; Length 354;
Best Local Similarity 24.1%; Pred. No. 1.7;
Matches 38; Conservative 17; Mismatches 63; Indels 40; Gaps 7;

QY 10 AAIVVGSALAG-VYPQWGGGNNHGG-----NSSGPDSTLSIYQGSAN 54
Db 200 AGDVNSGVGNAGDVNTGLNSGNINTGGFNPGLTGTFFSANTQAGFNS--GFENACTGN 257
QY 55 AALAQSDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFENNAITQWNAKNSDIT 114
Db 258 SGFGHNDPAGSGNGIQNSGFGNS-----GYVNTSTTSMFGNSGVLN 300
QY 115 VQYGGNNAALVNTASDSSVMVRQV---GFGNNATAN 149
Db 301 TG-YG--NSGFYNAAVNNTGIFVTGVMSGPFNFEGTGN 335

RESULT 8
US-09-730-374-3
; Sequence 3, Application US/09730374
; Patent No. US20010031261A1
; GENERAL INFORMATION:
; APPLICANT: Lust, John A.
; APPLICANT: Donovan, Kathleen A.
; TITLE OF INVENTION: USE OF GENETICALLY ENGINEERED ANTIBODIES
; FILE REFERENCE: 150.188US2
; CURRENT FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: US/09/730,374
; PRIOR FILING DATE: 2001-05-10
; PRIOR FILING DATE: 1999-06-04
; PRIOR APPLICATION NUMBER: 60/088,277
; PRIOR FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 249
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: A polypeptide encoded by an open reading frame of
; US-09-730-374-3

Query Match      11.3%; Score 87.5; DB 9; Length 249;
Best Local Similarity 27.3%; Pred. No. 1.6;
Matches 27; Conservative 9; Mismatches 32; Indels 31; Gaps 2;

QY 27 GGGGNHNGGNSGPD-----STLSIYQGSANAAL 57
Db 126 GGGSGGGGGGGSDIELTQSPSSFSVSLGDRVTITCKASEDIYNRLAWYQKPGNAPR 185
QY 58 ALQSDARKSETTITQ--SGYNGADYDQLVTRVVTHEMA 94
Db 186 LLISGATSLETGVPFRFSGSGKDYTLTSLQTEDVA 224

RESULT 9
US-10-704-206-3
; Sequence 3, Application US/10704206
; Publication No. US20040141982A1
; GENERAL INFORMATION:
; APPLICANT: Lust, John A.
; APPLICANT: Donovan, Kathleen A.
; TITLE OF INVENTION: USE OF GENETICALLY ENGINEERED ANTIBODIES TO CD38 TO TREAT MULTIPLE
; FILE REFERENCE: 150.188US2
; CURRENT FILING DATE: 2003-11-07
```

```
; OTHER INFORMATION: PPE
; NAME/KEY: misc.feature
; OTHER INFORMATION: GI|1781260
US-09-820-843A-21

Query Match      11.5%; Score 89; DB 10; Length 354;
Best Local Similarity 24.1%; Pred. No. 1.7;
Matches 38; Conservative 17; Mismatches 63; Indels 40; Gaps 7;

QY 10 AAIVVGSALAG-VYPQWGGGNNHGG-----NSSGPDSTLSIYQGSAN 54
Db 200 AGDVNSGVGNAGDVNTGLNSGNINTGGFNPGLTGTFFSANTQAGFNS--GFENACTGN 257
QY 55 AALAQSDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFENNAITQWNAKNSDIT 114
Db 258 SGFGHNDPAGSGNGIQNSGFGNS-----GYVNTSTTSMFGNSGVLN 300
QY 115 VQYGGNNAALVNTASDSSVMVRQV---GFGNNATAN 149
Db 301 TG-YG--NSGFYNAAVNNTGIFVTGVMSGPFNFEGTGN 335

RESULT 8
US-09-730-374-3
; Sequence 3, Application US/09730374
; Patent No. US20010031261A1
; GENERAL INFORMATION:
; APPLICANT: Lust, John A.
; APPLICANT: Donovan, Kathleen A.
; TITLE OF INVENTION: USE OF GENETICALLY ENGINEERED ANTIBODIES
; FILE REFERENCE: 150.188US2
; CURRENT FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: US/09/730,374
; PRIOR FILING DATE: 2001-05-10
; PRIOR FILING DATE: 1999-06-04
; PRIOR APPLICATION NUMBER: 60/088,277
; PRIOR FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 249
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: A polypeptide encoded by an open reading frame of
; US-09-730-374-3

Query Match      11.3%; Score 87.5; DB 9; Length 249;
Best Local Similarity 27.3%; Pred. No. 1.6;
Matches 27; Conservative 9; Mismatches 32; Indels 31; Gaps 2;

QY 27 GGGGNHNGGNSGPD-----STLSIYQGSANAAL 57
Db 126 GGGSGGGGGGGSDIELTQSPSSFSVSLGDRVTITCKASEDIYNRLAWYQKPGNAPR 185
QY 58 ALQSDARKSETTITQ--SGYNGADYDQLVTRVVTHEMA 94
Db 186 LLISGATSLETGVPFRFSGSGKDYTLTSLQTEDVA 224

RESULT 9
US-10-704-206-3
; Sequence 3, Application US/10704206
; Publication No. US20040141982A1
; GENERAL INFORMATION:
; APPLICANT: Lust, John A.
; APPLICANT: Donovan, Kathleen A.
; TITLE OF INVENTION: USE OF GENETICALLY ENGINEERED ANTIBODIES TO CD38 TO TREAT MULTIPLE
; FILE REFERENCE: 150.188US2
; CURRENT FILING DATE: 2003-11-07
```


Db 299 AGNGA-----SGNGHNYGNGGDDVITAPITGVNLISGFTLIGNSSSSVNT 352
Qy 56 ALALQSDARKSETTI-----TQSGYNGADYDQLVTRVVTHEMAHAFRNATIDWNA 108
Db 353 APTTSTVNDNDTIDNGNSGCGTSGGSGG-DLLNGAASNGER-----NYGNGNG 407
Qy 109 KNSDIT-----VGQVGGNNAALVNOTASDS 133
Db 408 DDVITAPITGVNFSGNSFSLIGNSSSS 437

RESULT 14

US-10-282-122A-64364
; Sequence 64364, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Chisen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELTRA 034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/205,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 64364
; LENGTH: 2204
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis

US-10-282-122A-64364

Query Match 11.0%; Score 85.5; DB 12; Length 2204;
Best Local Similarity 25.0%; Pred. No. 38;
Matches 34; Conservative 15; Mismatches 56; Indels 31; Gaps 4;
Qy 30 GNHNGGNSGPDST-----LSIYQGSANAALQSDARKSETTITQSGYNGADYDQLV 85
Db 1886 GGVNIGNNGIANTGTFIDGLANLGSYNTGLANLGD-----DNLGFGNAGSYNIGF 1937
Qy 86 TRVTHEMAHAFRNATIDWNAKNSDITVG-----QYGGNNAALVNOTASDS 133
Db 1938 ANFGSDNLGFPANTGSYNTGNTGNNIGVLTGNGQIGIGLSNGSNNIGLNSGSGN- 1996

Qy 134 SVMVRQVGFNNATAN 149
Db 1997 -----IGFFNSGTGN 2006

RESULT 15

US-09-880-748-905
; Sequence 905, Application US/09880748
; Publication No. US20030059937A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Antibodies that Immunospecifically Bind Blys
; FILE REFERENCE: PFS23
; CURRENT APPLICATION NUMBER: US/09/880,748
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 60/212,210
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: 60/240,816
; PRIOR FILING DATE: 2000-10-17
; PRIOR APPLICATION NUMBER: 60/276,248
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/277,379
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/293,499
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 3239
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 905
; LENGTH: 254
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-880-748-905

Query Match 11.0%; Score 85; DB 10; Length 254;
Best Local Similarity 23.0%; Pred. No. 2.9;
Matches 29; Conservative 11; Mismatches 42; Indels 44; Gaps 3;

Qy 13 VVSGSALAGVWPQNG-----GGNHNGGNSGPDSTLSI----- 47
Db 104 ILTYMGSAFDQWGRGTMTVSSGGGGGGGGGSALEI VNTQSPGTLISLSPGERA 163
Qy 48 -----YQYGSANAALQSDARKSETTITQ--SGYNGADYDQLVTRV 88
Db 164 TLSCRASQSI RSNYLAWYQKSGQAPRLLIYDVSSRATGIPDRFSGSGSGTDFLTISRL 223
Qy 89 VTHEMA 94
Db 224 EPEDFA 229

Search completed: August 2, 2004, 15:36:13
Job time : 36.8 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.
OM protein - protein search, using sw model
Run on: August 2, 2004, 14:48:33 ; Search time 167.9 Seconds
(without alignments)
877.809 Million cell updates/sec
Title: US-09-543-407-28
Perfect score: 775
Sequence: 1 MKLLKVAAPRAIVVSGSALA.....DSSVMVRQVFGNNATANQV 151
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 6019581 seqs, 976053577 residues
Total number of hits satisfying chosen parameters: 6019581
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :		Pending Patents AA Main:*	
1:	/cgn2_6/ptodata/2/paa/PCRTUS_COMB.pep.*	151	19
2:	/cgn2_6/ptodata/2/paa/US06_COMB.pep.*	151	19
3:	/cgn2_6/ptodata/2/paa/US07_COMB.pep.*	151	19
4:	/cgn2_6/ptodata/2/paa/US08_COMB.pep.*	151	6
5:	/cgn2_6/ptodata/2/paa/US09_COMB.pep.*	151	19
6:	/cgn2_6/ptodata/2/paa/US10_COMB.pep.*	151	19
7:	/cgn2_6/ptodata/2/paa/US11_COMB.pep.*	151	19
8:	/cgn2_6/ptodata/2/paa/US12_COMB.pep.*	151	19
9:	/cgn2_6/ptodata/2/paa/US13_COMB.pep.*	151	19
10:	/cgn2_6/ptodata/2/paa/US14_COMB.pep.*	151	19
11:	/cgn2_6/ptodata/2/paa/US15_COMB.pep.*	151	19
12:	/cgn2_6/ptodata/2/paa/US16_COMB.pep.*	151	19
13:	/cgn2_6/ptodata/2/paa/US17_COMB.pep.*	151	19
14:	/cgn2_6/ptodata/2/paa/US18_COMB.pep.*	151	19
15:	/cgn2_6/ptodata/2/paa/US19_COMB.pep.*	151	19
16:	/cgn2_6/ptodata/2/paa/US20_COMB.pep.*	151	19
17:	/cgn2_6/ptodata/2/paa/US21_COMB.pep.*	151	19
18:	/cgn2_6/ptodata/2/paa/US22_COMB.pep.*	151	19
19:	/cgn2_6/ptodata/2/paa/US23_COMB.pep.*	151	19
20:	/cgn2_6/ptodata/2/paa/US24_COMB.pep.*	151	19
21:	/cgn2_6/ptodata/2/paa/US25_COMB.pep.*	151	19
22:	/cgn2_6/ptodata/2/paa/US26_COMB.pep.*	151	19
23:	/cgn2_6/ptodata/2/paa/US27_COMB.pep.*	151	19
24:	/cgn2_6/ptodata/2/paa/US28_COMB.pep.*	151	19
25:	/cgn2_6/ptodata/2/paa/US29_COMB.pep.*	151	19
26:	/cgn2_6/ptodata/2/paa/US30_COMB.pep.*	151	19
27:	/cgn2_6/ptodata/2/paa/US31_COMB.pep.*	151	19
28:	/cgn2_6/ptodata/2/paa/US32_COMB.pep.*	151	19
29:	/cgn2_6/ptodata/2/paa/US33_COMB.pep.*	151	19
30:	/cgn2_6/ptodata/2/paa/US34_COMB.pep.*	151	19
31:	/cgn2_6/ptodata/2/paa/US35_COMB.pep.*	151	19
32:	/cgn2_6/ptodata/2/paa/US36_COMB.pep.*	151	19
33:	/cgn2_6/ptodata/2/paa/US37_COMB.pep.*	151	19

1	775	100.0	151	19	US-09-543-407-28	Sequence 28, Appl
2	712	91.9	151	19	US-09-543-407-22	Sequence 22, Appl
3	683	88.1	151	19	US-09-543-407-5	Sequence 5, Appl
4	678	87.5	151	6	US-08-233-642A-57	Sequence 57, Appl
5	662	85.4	151	19	US-09-543-407-20	Sequence 20, Appl
6	617	79.6	151	19	US-09-543-407-26	Sequence 26, Appl
7	613	79.1	151	19	US-09-543-407-30	Sequence 30, Appl
8	605	78.1	151	19	US-09-543-407-12	Sequence 12, Appl
9	603	77.8	151	19	US-09-543-407-14	Sequence 14, Appl
10	601	77.5	151	19	US-09-543-407-18	Sequence 18, Appl
11	600	77.4	151	19	US-09-543-407-24	Sequence 24, Appl
12	596	76.9	131	19	US-09-543-407-31	Sequence 31, Appl
13	568	73.3	151	19	US-09-543-407-16	Sequence 16, Appl
14	509	65.7	151	19	US-09-543-407-7	Sequence 7, Appl
15	506	65.3	151	13	US-08-978-878-4	Sequence 4, Appl
16	506	65.3	151	21	US-09-741-873B-4	Sequence 4, Appl
17	504	65.0	151	33	US-60-352-946-2	Sequence 2, Appl
18	504	65.0	151	33	US-60-444-371-2	Sequence 2, Appl
19	498	64.3	120	6	US-08-233-642A-55	Sequence 55, Appl
20	464	59.9	109	19	US-09-543-407-34	Sequence 34, Appl
21	462	59.6	158	16	US-09-252-691-5834	Sequence 5834, Ap
22	462	59.6	158	16	US-09-252-691C-5833	Sequence 5833, Ap
23	462	59.6	158	30	US-10-417-886-5834	Sequence 5834, Ap
24	428	55.2	131	13	US-08-978-878-2	Sequence 2, Appl
25	428	55.2	131	21	US-09-741-873B-2	Sequence 35, Appl
26	324	41.8	109	19	US-09-543-407-37	Sequence 37, Appl
27	257	33.2	68	19	US-09-543-407-39	Sequence 39, Appl
28	181	23.4	48	19	US-09-543-407-32	Sequence 32, Appl
29	160.5	20.7	70	19	US-09-543-407-33	Sequence 5833, Ap
30	102.5	13.2	186	16	US-09-252-691-5833	Sequence 5833, Ap
31	102.5	13.2	186	16	US-09-252-691C-5833	Sequence 5833, Ap
32	102.5	13.2	186	30	US-10-417-886-5833	Sequence 18734, A
33	95	12.3	400	21	US-09-733-089-18734	Sequence 18734, A
34	95	12.3	400	23	US-09-816-660-18734	Sequence 28424, A
35	95	12.3	400	27	US-10-155-881-28424	Sequence 186417, A
36	95	12.3	400	30	US-10-437-963-186417	Sequence 17264, A
37	95	12.3	400	30	US-10-438-246-17264	Sequence 18735, A
38	95	12.3	1060	21	US-09-733-089-18735	Sequence 18735, A
39	95	12.3	1060	23	US-09-816-660-18735	Sequence 6, Appl
40	94.5	12.2	151	19	US-09-543-407-6	Sequence 41783, A
41	92.5	11.9	644	22	US-09-791-537-108562	Sequence 108562, A
42	92	11.9	598	22	US-09-791-537-108562	Sequence 50616, A
43	92	11.9	1074	28	US-10-282-122A-50616	Sequence 23883, A
44	91.5	11.8	478	22	US-09-791-537-23883	Sequence 29231, A
45	91.5	11.8	602	22	US-09-791-537-29231	

ALIGNMENTS

RESULT 1
US-09-543-407-28
; Sequence 28, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Ray, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	DB	ID	Description

US-09-543-407-28

```
Query Match      100.0%; Score 775; DB 19; Length 151;
Best Local Similarity 100.0%; Pred. NO. 1.2e-74;
Matches 151; Conservative 0; Mismatches 0; Indels 0
```

1	MKLLKVAAPAAI	VVSGSALAGV	PQWGGGNGHNGGNS	GGPDSTLSI	YQVGSNARALAQ	60
2						
3	1	MKLLKVAAPAAI	VVSGSALAGV	PQWGGGNGHNGGNS	GGPDSTLSI	YQVGSNARALAQ
4						
5	61	SDARKSETTITQSGY	GNAGADYQLVTR	VVVTHEMAHAFNNATIDQWNAKNSDITVQYGG	120	
6						
7	61	SDARKSETTITQSGY	GNAGADYQLVTR	VVVTHEMAHAFNNATIDQWNAKNSDITVQYGG	120	
8						
9	121	NNAALVNQTASDSSVM	VRQVGFNNATANQY	151		
10						
11	121	NNAALVNQTASDSSVM	VRQVGFNNATANQY	151		

RESULT 2

US-09-543-407-22

```

: Sequence 22, Application US/09543407
: GENERAL INFORMATION:
: APPLICANT: White, Aaron P.
: APPLICANT: Dorian, James L.
: APPLICANT: Collinson, S. Karen
: APPLICANT: Kay, William W.
: TITLE OF INVENTION: BACTERIAL FIMBRIAL SVSTM FOR
: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
: FILE REFERENCE: 920043.406
: CURRENT APPLICATION NUMBER: US/09/543,407
: CURRENT FILING DATE: 2000-04-05
: NUMBER OF SEQ ID NOS: 59
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 22
: LENGTH: 151
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
: OTHER INFORMATION: sequence containing the replacement fragment
: OTHER INFORMATION: encoding P73 from GPe3 of Leishmania major.
: US-09-543-407-22

```

Query Match 91.9%; Score 712; DB 19; Length 151;
Best Local Similarity 91.1%; Pred. No. 7.2e-68;
Matches 144; Conservative 0; Mismatches 0; Indels 14; Gaps 2;

```

1  MKLLKVAARFAAIIVSGSALAGVVPQWGGGHHNGGNSGPGDSTLSIYQVGSNAALALQ  60
1  MKLLKVAARFAAIIVSGSALAGVVPQWGGGHHNGGNSGPGDSTLSIYQVGSNAALALQ  60
61  SDARKSETTITOSGYNGAD-----YDQIVTEVWTHMAHAFRNATIDQNAKNSDI  113
61  SDARKSETTITOSGYNGADVGQADNYDQIVTEVWTHMAH-----DOWNAKNSDI  113

```

114 TVGQYGGNNAALVNQTASDSSVMVRQVGFNNATANQY 151
114 TVGQYGGNNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 3

US-09-543-407-5

Sequence 5, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Dotan, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIBRILAR SYSTEM FOR
PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
FILE REFERENCE: 920043.496
CURRENT APPLICATION NUMBER: US/09/543,407

; CURRENT FILING DATE: 2000-04-05

```

; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 151

```

```
Query Match      88.1%; Score 683; DB 19; Length 151;
Best Local Similarity 90.1%; Pred. No. 9.5e-65;
Matches 136; Conservative 1; Mismatches 14; Indels 0; Caps 0;
```

1	QY	1	KKLKVAAFAAIUVVSGSNAIAGVDPQWCGGNGHNGCGNSGPDSTLSIYQVGSNAALALQ	60
1	Db	1	KKLKVAAFAAIUVVSGSNAIAGVDPQWCGGNGHNGCGNSGPDSTLSIYQVGSNAALALQ	60
61	QY	61	SDARKSETTITQSGYNGGADYQDLVTRVWTHMAHAFRNATTDQWNAKNSDITVGYQYG	120
61	Db	61	SDARKSETTITQSGYNGGADYQDQADNSTLTQNGFRNATTDQWNAKNSDITVGYQYG	120
121	QY	121	NNAALVNQATSDSSVMVROVGFNGGNATANCY	151
121	Db	121	NNAALVNQATSDSSVMVROVGFNGGNATANCY	151

RESULT 4

US-08-233-642A-57

```

1 Sequence 57, Application US/08233642A
2
3 GENERAL INFORMATION:
4
5 APPLICANT: Kay, William W.
6 APPLICANT: Collinson, S. Karen
7 APPLICANT: Clouthier, Sharon C.
8 APPLICANT: Doran, James L.
9
10 TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
11 BASED VACCINES
12
13 NUMBER OF SEQUENCES: 58
14 NUMBER OF SEQUENCES: -
15 CORRESPONDENCE ADDRESS:
16 ADDRESSEE: Seed and Berry
17 STREET: 6300 Columbia Center, 701 Fifth Avenue
18 CITY: Seattle
19 STATE: Washington
20 COUNTRY: U.S.A.
21 ZIP: 98104-7092
22
23 COMPUTER READABLE FORM:
24 MEDIUM TYPE: Floppy disk
25 COMPUTER: IBM PC compatible
26 OPERATING SYSTEM: PC-DOS/MS-DOS
27 SOFTWARE: Patent in Release #1.0, Version #1.25
28 CURRENT APPLICATION DATA:
29 APPLICATION NUMBER: US/08/233,642A
30 FILING DATE: 26-APR-1994
31 CLASSIFICATION: 424
32 ATTORNEY/AGENT INFORMATION:
33 NAME: King, Joshua
34 REGISTRATION NUMBER: 35,570
35 REFERENCE/DOCKET NUMBER: 920043.403C3
36 TELECOMMUNICATION INFORMATION:
37 TELEPHONE: (206) 622-4900
38 TELEFAX: (206) 682-6031
39 TELEX: 3723836 SEEDANBERRY
40 INFORMATION FOR SEQ ID NO: 57:
41 SEQUENCE CHARACTERISTICS:
42 LENGTH: 151 amino acids
43 TYPE: amino acid
44 TOPOLOGY: linear
45 MOLECULE TYPE: protein
46
47 US-08-233-642A-57

```

Query Match 87.5%; Score 678; DB 6; Length 151;
Best Local Similarity 89.4%; Pred. No. 3.3e-64;
Matches 135; Conservative 1; Mismatches 15; Indels

FILE REFERENCE: 920043.408
CURRENT APPLICATION NUMBER: US/09/543,407

```
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALQ 60
QY 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAFRNNATIDQWNAKNSDITVGOYGG 120
DB 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAFRNNATIDQWNAKNSDITVGOYGG 120
QY 121 NNALVNTASDSSVMVROVGFNNATANQY 151
DB 121 NNALVNTASDSSVMVROVGFNNATANQY 151

RESULT 5
US-09-543-407-20
; Sequence 20 Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-20
Query Match 85.4%; Score 662; DB 19; Length 151;
Best Local Similarity 81.9%; Pred. No. 1.7e-62;
Matches 136; Conservative 0; Mismatches 0; Indels 30; Gaps 2;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALQ 60
QY 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAFRNNATIDQ 105
DB 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAFRNNATIDQ 105
QY 106 WNAKNSDITVGOYGGNNAALVNQTASDSSVMVROVGFNNATANQY 151
DB 106 WNAKNSDITVGOYGGNNAALVNQTASDSSVMVROVGFNNATANQY 151

RESULT 6
US-09-543-407-26
; Sequence 26 Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-26
Query Match 79.1%; Score 613; DB 19; Length 151;
Best Local Similarity 74.6%; Pred. No. 3.2e-57;
Matches 129; Conservative 0; Mismatches 0; Indels 44; Gaps 2;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALQ 60
QY 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAFRNNATIDQ 98
DB 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAFRNNATIDQ 98
QY 99 NNATIDQWNAKNSDITVGOYGGNNAALVNQTASDSSVMVROVGFNNATANQY 151
DB 119 NNATIDQWNAKNSDITVGOYGGNNAALVNQTASDSSVMVROVGFNNATANQY 151

RESULT 7
US-09-543-407-30
; Sequence 30 Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-30
Query Match 79.1%; Score 613; DB 19; Length 151;
Best Local Similarity 74.6%; Pred. No. 3.2e-57;
Matches 129; Conservative 0; Mismatches 0; Indels 44; Gaps 2;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALQ 60
QY 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAFRNNATIDQ 98
DB 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAFRNNATIDQ 98
QY 99 NNATIDQWNAKNSDITVGOYGGNNAALVNQTASDSSVMVROVGFNNATANQY 151
DB 119 NNATIDQWNAKNSDITVGOYGGNNAALVNQTASDSSVMVROVGFNNATANQY 151

RESULT 8
US-09-543-407-12
; Sequence 12 Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
```

```
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-26
Query Match 79.6%; Score 617; DB 19; Length 151;
Best Local Similarity 73.6%; Pred. No. 1.2e-57;
Matches 128; Conservative 0; Mismatches 0; Indels 46; Gaps 2;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALQ 57
QY 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAFRNNATIDQ 97
DB 58 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAFRNNATIDQ 97
QY 98 RNNATIDQWNAKNSDITVGOYGGNNAALVNQTASDSSVMVROVGFNNATANQY 151
DB 98 RNNATIDQWNAKNSDITVGOYGGNNAALVNQTASDSSVMVROVGFNNATANQY 151

RESULT 7
US-09-543-407-30
; Sequence 30 Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-30
Query Match 79.1%; Score 613; DB 19; Length 151;
Best Local Similarity 74.6%; Pred. No. 3.2e-57;
Matches 129; Conservative 0; Mismatches 0; Indels 44; Gaps 2;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALQ 60
QY 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAFRNNATIDQ 98
DB 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAFRNNATIDQ 98
QY 99 NNATIDQWNAKNSDITVGOYGGNNAALVNQTASDSSVMVROVGFNNATANQY 151
DB 119 NNATIDQWNAKNSDITVGOYGGNNAALVNQTASDSSVMVROVGFNNATANQY 151

RESULT 8
US-09-543-407-12
; Sequence 12 Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
```

```

; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-12

Query Match 78.1%; Score 605; DB 19; Length 151;
Best Local Similarity 80.1%; Pred. No. 2.3e-56;
Matches 121; Conservative 5; Mismatches 25; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPWGCGGNHNGGSSGPDSTLSIYQYGSANAALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPWGCGGNHNGGSSGPDSTLSIYQYGSANAALQ 60
QY 61 SDARKSETTITQSGYNGADVDQLVTRVVTHEMAHAFRNATIDQWNAKNSDITVGYGG 120
DB 61 SDARKSETTITQSGYNGADVDQLVTRVVTHEMAHAFRNATIDQWNAKNSDITVGYGG 120
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
DB 121 NNAALVNQDQLVTRVVTHEMAHANATANQY 151

RESULT 9
US-09-543-407-14
; Sequence 14, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-14

Query Match 77.8%; Score 603; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 3.8e-56;
Matches 122; Conservative 4; Mismatches 25; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPWGCGGNHNGGSSGPDSTLSIYQYGSANAALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPWGCGGNHNGGSSGPDSTLSIYQYGSANAALQ 60
QY 61 SDARKSETTITQSGYNGADVDQLVTRVVTHEMAHAFRNATIDQWNAKNSDITVGYGG 120
DB 61 SDARKSETTITQSGYNGADVDQLVTRVVTHEMAHAFRNATIDQWNAKNSDITVGYGG 120

Query Match 77.5%; Score 601; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 6.3e-56;
Matches 122; Conservative 4; Mismatches 25; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPWGCGGNHNGGSSGPDSTLSIYQYGSANAALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPWGCGGNHNGGSSGPDYDQLVTRVVTHEMAHALQ 60
QY 61 SDARKSETTITQSGYNGADVDQLVTRVVTHEMAHAFRNATIDQWNAKNSDITVGYGG 120
DB 61 SDARKSETTITQSGYNGADVDQLVTRVVTHEMAHAFRNATIDQWNAKNSDITVGYGG 120
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 11
US-09-543-407-24
; Sequence 24, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-24

Query Match 77.4%; Score 600; DB 19; Length 151;

```

```

QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
DB 121 LVTRVVTHEMAHASVMVRQVGFNNATANQY 151

RESULT 10
US-09-543-407-18
; Sequence 18, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-18

Query Match 77.5%; Score 601; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 6.3e-56;
Matches 122; Conservative 4; Mismatches 25; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPWGCGGNHNGGSSGPDSTLSIYQYGSANAALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPWGCGGNHNGGSSGPDYDQLVTRVVTHEMAHALQ 60
QY 61 SDARKSETTITQSGYNGADVDQLVTRVVTHEMAHAFRNATIDQWNAKNSDITVGYGG 120
DB 61 SDARKSETTITQSGYNGADVDQLVTRVVTHEMAHAFRNATIDQWNAKNSDITVGYGG 120
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 11
US-09-543-407-24
; Sequence 24, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-24

Query Match 77.4%; Score 600; DB 19; Length 151;

```

Best Local Similarity 81.5%; Pred. No. 8.1e-56;
Matches 123; Conservative 3; Mismatches 25; Indels 0; Gaps 0;
QY 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
Db 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFRNATIDOWNAKNSDITVGYGG 120
Db 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFRNATIDOWNAKNSDITVGYGG 120

RESULT 12

US-09-543-407-31
; Sequence 31, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543.407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Salmonella enteritidis
US-09-543-407-31

Query Match 76.9%; Score 596; DB 19; Length 131;
Best Local Similarity 88.5%; Pred. No. 1.8e-55;
Matches 116; Conservative 1; Mismatches 14; Indels 0; Gaps 0;
QY 21 GWVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGAD 80
Db 1 GWVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGAD 60
QY 81 YDQLVTRVVTHEMAHAFRNATIDOWNAKNSDITVGYGNNALVNQTASDSSVMVRQV 140
Db 61 VGQADNSTIETQNGFRNATIDOWNAKNSDITVGYGNNALVNQTASDSSVMVRQV 120
QY 141 GFGNNATANQY 151
Db 121 GFGNNATANQY 131

RESULT 13

US-09-543-407-16
; Sequence 16, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543.407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence

FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-16

Query Match 73.3%; Score 568; DB 19; Length 151;
Best Local Similarity 80.1%; Pred. No. 2.2e-52;
Matches 121; Conservative 4; Mismatches 26; Indels 0; Gaps 0;
QY 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
Db 1 MKLLKVAAPAAIIVSGSALAGVYDQLVTRVVTHEMAHAFRNATIDOWNAKNSDITVGYGG 60
QY 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFRNATIDOWNAKNSDITVGYGG 120
Db 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFRNATIDOWNAKNSDITVGYGG 120
QY 121 NNAALVNQTASDSSVMVRQVFGNNATANQY 151
Db 121 NNAALVNQTASDSSVMVRQVFGNNATANQY 151

RESULT 14

US-09-543-407-7
; Sequence 7, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543.407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-543-407-7

Query Match 65.7%; Score 509; DB 19; Length 151;
Best Local Similarity 68.2%; Pred. No. 5e-46;
Matches 103; Conservative 15; Mismatches 33; Indels 0; Gaps 0;
QY 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
Db 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFRNATIDOWNAKNSDITVGYGG 120
Db 61 TDARNSDLTITQGGGNGADYDQGGSDSSIDLTRQFGNSATLDQNGKNSMTVKQFGG 120
QY 121 NNAALVNQTASDSSVMVRQVFGNNATANQY 151
Db 121 GNGAAVDQTASNSVNVTVQVFGNNATAHOY 151

RESULT 15

US-08-978-878-4
; Sequence 4, Application US/08978878
; GENERAL INFORMATION:
; APPLICANT: NORMARK, Staffan
; APPLICANT: OLSEN, Arne
; TITLE OF INVENTION: FIBRONECTIN BINDING PROTEIN AS WELL AS ITS PREPARATION
; FILE REFERENCE: 012889-C81
; CURRENT APPLICATION NUMBER: US/08/978.878
; CURRENT FILING DATE: 1997-11-26
; EARLIER APPLICATION NUMBER: SE 8801723-1
; EARLIER FILING DATE: 1988-05-06

Tue Aug 3 10:54:44 2004

```

; EARLIER APPLICATION NUMBER: US 07/347,189
; EARLIER FILING DATE: 1989-05-04
; EARLIER APPLICATION NUMBER: US 07/789,437
; EARLIER FILING DATE: 1991-11-06
; EARLIER APPLICATION NUMBER: US 07/970,846
; EARLIER FILING DATE: 1992-11-03
; EARLIER APPLICATION NUMBER: US 08/187,865
; EARLIER FILING DATE: 1994-01-28
; EARLIER APPLICATION NUMBER: US 08/318,519
; EARLIER FILING DATE: 1994-10-05
; EARLIER APPLICATION NUMBER: US 08/495,959
; EARLIER FILING DATE: 1995-06-28
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-08-978-878-4

Query Match      65.3%; Score 506; DB 13; Length 151;
Best Local Similarity 67.5%; Pred. No. 1e-45;
Matches 102; Conservative 16; Mismatches 33; Indels 0; Gaps 0;

QY 1 M K L L K V A A P A I V V S G S A L A G V P Q W G G G N H G G N S S G P D S T L S I Y Q G S A N A A L A Q 60
Db 1 M K L L K V A A I A I V F S G S A G A G V P Q Y G G G N H G G N S G P N S E L N I Y Q G G N S A L A Q 60

QY 61 S D A R K S E T T I T Q S G Y G N G A D Y D Q L V T R V V T H E M A H A F P N N A T I D Q W N A K N S D I T V Q Y G G 120
Db 61 T D A R N S D L T I T Q H G G N G A D V G Q G S D S I D L T Q R G C N S A T L D Q W N G K N S E M T V K Q F G G 120

QY 121 N N A A I V N O T A S D S Y M V R Q V G F G N N A T A N O Y 151
Db 121 G N G A A V D O T A S N S S Y N V T Q V G F G N N A T A H Q Y 151

```

Search completed: August 2, 2004, 15:26:46
Job time : 167.9 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:49:38 ; Search time 17.8 seconds
(without alignments)
888.146 Million cell updates/sec

Title: US-09-543-407-28

Perfect score: 775
Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVQVGFNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 601315 seqs, 104695340 residues

Total number of hits satisfying chosen parameters: 601315

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents AA New:*

- 1: /cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pcp:**
- 2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pcp:**
- 3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pcp:**
- 4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pcp:**
- 5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pcp:**
- 6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pcp:**
- 7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pcp:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	506	65.3	151	5	US-09-741-873C-4
2	428	55.2	131	5	US-09-741-873C-2
3	89.5	11.5	388	5	US-09-248-796A-17306
4	87.5	11.3	619	1	PCT-US04-10229-38
5	87.5	11.3	1905	1	PCT-US04-09388-9
6	84.5	10.9	234	6	US-10-767-701-45603
7	84.5	10.9	511	6	US-10-425-115-320950
8	82	10.6	276	6	US-10-425-115-339993
9	81	10.5	147	6	US-10-425-115-193207
10	81	10.5	224	6	US-10-805-394-4894
11	81	10.5	457	5	US-09-248-796A-26659
12	81	10.5	568	1	PCT-US04-21492-458
13	80	10.3	1627	6	US-10-170-205E-16659
14	79	10.2	295	6	US-10-425-115-312468
15	79	10.2	1010	1	PCT-US04-21492-402
16	78	10.1	892	5	US-09-952-267B-5
17	78	10.1	892	6	US-10-872-768-5
18	78	10.1	892	6	US-10-872-769-5
19	77.5	10.0	246	6	US-10-854-439-511
20	77.5	10.0	415	5	US-09-490-953-13
21	77.5	10.0	520	6	US-10-479-638-21
22	77	9.9	1160	6	US-10-603-150-2
23	76.5	9.9	132	6	US-10-425-115-351875
24	76.5	9.9	248	6	US-10-170-205E-3429
25	76.5	9.9	283	6	US-10-854-439-1222
26	76.5	9.9	685	5	US-09-248-796A-15216

27 76.5 9.9 873 5 US-09-952-267B-13 Sequence 13, Appl
28 76.5 9.9 873 6 US-10-872-768-13 Sequence 13, Appl
29 76.5 9.9 873 6 US-10-872-769-13 Sequence 13, Appl
30 76.5 9.9 1058 6 US-10-796-907-72 Sequence 72, Appl
31 76 9.8 127 6 US-10-425-115-274032 Sequence 274032, Appl
32 76 9.8 201 6 US-10-425-115-309662 Sequence 309662, Appl
33 76 9.8 279 6 US-10-425-115-343835 Sequence 343835, Appl
34 76 9.8 598 7 US-10-425-115-221471 Sequence 221471, Appl
35 76 9.8 688 7 US-60-581-351-11388 Sequence 11388, A
36 76 9.8 753 7 US-60-581-351-12224 Sequence 12224, A
37 76 9.8 832 7 US-10-425-115-287789 Sequence 287789, Appl
38 76 9.8 3132 7 US-60-556-841-1245 Sequence 1245, Appl
39 75.5 9.7 400 6 US-10-490-953-13 Sequence 13, Appl
40 75.5 9.7 400 6 US-10-490-953-14 Sequence 14, Appl
41 75.5 9.7 400 6 US-10-490-953-20 Sequence 20, Appl
42 75.5 9.7 400 6 US-10-723-981-14 Sequence 14, Appl
43 75.5 9.7 400 6 US-10-723-981-15 Sequence 15, Appl
44 75.5 9.7 424 6 US-10-045-674A-591 Sequence 591, Appl
45 75.5 9.7 529 6 US-10-425-115-340009 Sequence 340009, Appl

ALIGNMENTS

RESULT 1

US-09-741-873C-4
; Sequence 4; Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873C-4

Query Match 65.3%; Score 506; DB 5; Length 151;
Best Local Similarity 67.5%; Pred. No. 6.1e-38;
Matches 102; Conservative 16; Mismatches 33; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQGGNGGNGSSGDPSTLSIYQYGSANAALAQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQYGGNGGNGGNGSPNSLNIYQYGGNSALALQ 60
QY 61 SPARKSETTITQSGYNGADYDQLVTRVVTHEMAFAFRNNATIDOWNAKNSDITVQYGG 120
DB 61 TARNESDLTITQGGNGGADVCGGSDSSIDYTORGFNGSATIDOWNKNSENTVQKQFG 120
QY 121 NNAALVNOTASDSSVMVQVGFNNATANQY 151
DB 121 GNGAAVDQTASNSSVNVTVQVGFNNATAHQY 151

```

RESULT 2
US-09-741-873C-2
; Sequence 2, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-08-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Escherichia coli
; US-09-741-873C-2

Query Match      55.2%; Score 428; DB 5; Length 131;
Best Local Similarity 64.1%; Pred. No. 5e-31;
Matches 84; Conservative 15; Mismatches 32; Indels 0; Gaps 0;

QY 21 GVVPQGGGNGHNGSGPDSLIIYQYGSANAALQSDARKSETTITQSGYNGAD 80
DB 1 GVVPQGGGNGHNGSGPDSLIIYQYGSANAALQSDARKSETTITQSGYNGAD 60
QY 81 YDLVTRVVTHEMAHAFRNATIDOWNAKNSDITVQYGGNNAALVNQTSADSSVMVQV 140
DB 61 VQGGSDSSIDITQRFNGSALTDQNGKNSMTVKQFGGNGAAGVDTASNSVNVTVQV 120
QY 141 GFGNNTANQY 151
DB 121 GFGNNTANQY 131

RESULT 3
US-09-248-796A-17306
; Sequence 17306, Application US/09248796A
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 17306
; LENGTH: 388
; TYPE: PRT
; ORGANISM: Candida albicans
; US-09-248-796A-17306

Query Match      11.5%; Score 89.5; DB 5; Length 388;
Best Local Similarity 22.1%; Pred. No. 3 6;
Matches 33; Conservative 17; Mismatches 44; Indels 55; Gaps 4;

```

```

QY 30 GNHNGGNSGPDSTLSIYQYGSANAALQSDARKSETTITQSGYNGADYDLVTRVY 89
DB 132 GNVVGSNSYSPDT-----YGSAITGLGQEKTPAVTGIHSGKIG-AAAYPELT----- 180
QY 90 THEMAHAFRNATIDOWNAKNSDITVQYGGNNAALVNQTS----- 131
DB 181 -----NAGNTGLAKGTAPATSTATSYGESPSADYSKSGATGVVPATVYN 223
QY 132 -----DSSVMVRQVGFNNATANQY 151
DB 224 TSCAPTGSINTAGVGGAGFGDNSNTSSY 252

RESULT 4
PCT-US04-10229-38
; Sequence 38, Application PC/TUS0410229
; GENERAL INFORMATION:
; APPLICANT: Diversa Corporation
; APPLICANT: Kerovuo, Janne
; APPLICANT: Solbak, Arne
; APPLICANT: Gray, Kevin
; APPLICANT: McCann, Ryan
; APPLICANT: Purohit, Shalaka
; APPLICANT: Gerendash, Joel
; APPLICANT: Janssen, Giselle
; APPLICANT: Dahod, Samun
; TITLE OF INVENTION: PECTATE LYASES, NUCLEIC ACIDS ENCODING THEM
; FILE REFERENCE: 564462009640
; CURRENT APPLICATION NUMBER: PCT/US04/10229
; CURRENT FILING DATE: 2003-04-03
; PRIOR APPLICATION NUMBER: 60/460,842
; PRIOR FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: 60/484,798
; PRIOR FILING DATE: 2003-07-03
; NUMBER OF SEQ ID NOS: 134
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 38
; LENGTH: 619
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)...(35)
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (36)...(387)
; OTHER INFORMATION: Catalytic domain
; PCT-US04-10229-38

Query Match      11.3%; Score 87.5; DB 1; Length 619;
Best Local Similarity 26.9%; Pred. No. 9.4; Indels 45; Gaps 4;
Matches 36; Conservative 10; Mismatches 43;

QY 3 LLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDS-TLSIYQYGSANAALQLOS 61
DB 18 LTAIAAVALMAGTSFAAAT---GGFSTTDGGNVSGARSFTASTYQQINTIIANAKLD 73
QY 62 DARKSETTITQSGYNGADYDLVTRVVTHEMAHAFRNATIDOWNAKNSDITVQYGCN 121
DB 74 DAGKVT-----GGAYPLIIT-----YTGN 93
QY 122 NAALVNQTSADSSV 135
DB 94 EDSLINQMIKDHV 107

RESULT 5
PCT-US04-09388-9
; Sequence 9, Application PC/TUS0409388

```



```
; GENERAL INFORMATION:
; APPLICANT: INCYTE CORPORATION; MARQUIS, Joseph P.;
; APPLICANT: TRAN, Uyen K.; YANG, Yonghong G.;
; APPLICANT: TANG, Y. Tom; CHAWLA, Narinder K.;
; APPLICANT: MURAGE, Jaji; WANG, Jonathan T.;
; APPLICANT: ELLIOTT, Vicki S.; CHEN, David;
; APPLICANT: YUE, Henry; AZIMZAI, Yalda;
; APPLICANT: ISON, Craig H.; KHARE, Reena;
; APPLICANT: JIN, Pei; RAMKUNAR, Javalaxmi;
; APPLICANT: FAVERO, Kristin D.; RICHARDSON, Thomas W.;
; APPLICANT: HAFALIA, April J.A.; BAUGHN, Mariah R.;
; APPLICANT: BECHA, Shanva D.; WILSON, Amy D.
; TITLE OF INVENTION: PROTEINS ASSOCIATED WITH CELL GROWTH, DIFFERENTIATION, AND DEATH
; FILE REFERENCE: PF-1531 PCT
; CURRENT APPLICATION NUMBER: PCT/US04/09388
; CURRENT FILING DATE: 2004-04-01
; PRIOR APPLICATION NUMBER: US 60/457,403
; PRIOR FILING DATE: 2003-03-24
; PRIOR APPLICATION NUMBER: US 60/465,568
; PRIOR FILING DATE: 2003-04-24
; PRIOR APPLICATION NUMBER: US 60/476,135
; PRIOR FILING DATE: 2003-06-04
; PRIOR APPLICATION NUMBER: US 60/476,583
; PRIOR FILING DATE: 2003-06-06
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PERL Program
; SEQ ID NO 9
; LENGTH: 1905
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No: 7525307CD1
PCT-US04-09388-9

Query Match 11.3%; Score 87.5; DB 1; Length 1905;
Best Local Similarity 28.6%; Pred. No. 35;
Matches 36; Conservative 16; Mismatches 39; Indels 35; Gaps 6;

Qy 10 AAIIVGSGALAGVVPQGG-GNHNGGNSGPDSTLSIYQGSANAALQSDARKSET 68
Db 1031 STLVDNGTAAGKPPSSGSGWDHPA-----BPPVAFGRAGAPVAASALCKPASKS-- 1081
Qy 69 TITQSGYNGADYDQLVTRVVTTHMAHAFRNATIDQWNAKNSDITVQYGGNNAALVNQ 128
Db 1082 --MQEGWSSGD-----EM-----NLSTSQWEDEGDV-----WNAASQES 1116

Qy 129 TASDSS 134
Db 1117 TSSCSS 1122

RESULT 6
US-10-767-701-45603
; Sequence 45603, Application US/10767701
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(5335)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 45603
; LENGTH: 234
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(234)
; OTHER INFORMATION: unsure at all Xaa locations
```

```
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-C82834_1.pep
US-10-767-701-45603

Query Match 10.9%; Score 84.5; DB 6; Length 234;
Best Local Similarity 24.2%; Pred. No. 5.5; Indels 65; Gaps 6;
Matches 39; Conservative 11; Mismatches 46;

Qy 1 MKLLKVAFAAIVV--SGSALAGVVPQW---GGGGNHNGGNN-----SSGPDSTL 45
Db 1 MATTKLAALCFIVLLGIGGANAARVARYVSAGGGGGGGGGRWEGGASRWGSGSGSGC 60
Qy 46 SIY-----OYGSANAALALQSDARKSETTITQSGYNGADYD 82
Db 61 GXYGEAGSGCXAYAQGGGGGGGGGGGSG-----SGYSGSGYG 102
Qy 83 QLVTRVVTTHMAHAFRNATIDQWNAKNSDITVQYGGNNA 123
Db 103 Q-----AGSGSGNGGAYAQGGAGGGGGGGGGYGGSGS 134

RESULT 7
US-10-425-115-320950
; Sequence 320950, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 320950
; LENGTH: 511
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(511)
; OTHER INFORMATION: unsure at all Xaa locations
; OTHER INFORMATION: Clone ID: MRT4577_55769C.1.pep
US-10-425-115-320950

Query Match 10.9%; Score 84.5; DB 6; Length 511;
Best Local Similarity 23.3%; Pred. No. 14;
Matches 44; Conservative 22; Mismatches 64; Indels 59; Gaps 8;

Qy 1 MKLLKVAFAAIVWS-----GSALAGVVPQWGG-----GNHNGGNSGSPD 42
Db 133 LRLLEERAQEAIALASCRDVLRSAGFFEDAWKVPISGDEGVYVAANYALGRLGGDPN 192
Qy 43 STLSTIYQGSANAALALQSDA-----RKSETTITQ-----SGYNGADYDQLVTR 87
Db 193 KTVGIIELGASQAQLTFVSDEVLPPKLSYNYTFGETTYTLTNSFLNFGQNAAQDSF--- 249
Qy 88 VVTHEVAH---AFRNATID---QWNAKNSDITVQYGGNNAALVNQTSADSSVMVQV 140
Db 250 ---HEMLRSGFKNGTLADPCAPRGYSRNEEMLRMSGASRSTLENQ----- 294
Qy 141 GFQNNATAN 149
Db 295 -YVNNGTGN 302

RESULT 8
US-10-425-115-339993
; Sequence 339993, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
```

APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants
FILE REFERENCE: 38-21(53222)B
CURRENT APPLICATION NUMBER: US/10/425,115
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 369326
SEQ ID NO 339993
LENGTH: 276
TYPE: PRT
ORGANISM: Zea mays
FEATURE:
NAME/KEY: unsure
LOCATION: (1)..(276)
OTHER INFORMATION: unsure at all Xaa locations
FEATURE:
OTHER INFORMATION: Clone ID: MRT4577_73240C.1.pep
US-10-425-115-339993

Query Match 10.6%; Score 82; DB 6; Length 276;
Best Local Similarity 27.1%; Pred. No. 11;
Matches 29; Conservative 17; Mismatches 33; Indels 28; Gaps 5;

QY 12 IVVSGSALAGVVPQWGGGNGHNGSSGPDSTLSIYQYGSANAALALQSDARKSETTIT 71
DB 10 LLVLGVSQGGTAQGGG-----GGGLTRG--SFPKGFVFGTA-----SSAYQYEGAVK 56

QY 72 QSGYNGGADYDQLVTRVVTHEMAHAFRNATIDQWNAKNSDITVQY 118
DB 57 EDGRG-----XTIWDKFAHTFGKVAIDL-----SNADVAVDQY 88

RESULT 9
US-10-425-115-193207
Sequence 193207, Application US/10425115
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J.
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants
FILE REFERENCE: 38-21(53222)B
CURRENT APPLICATION NUMBER: US/10/425,115
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 369326
SEQ ID NO 193207
LENGTH: 147
TYPE: PRT
ORGANISM: Zea mays
FEATURE:
OTHER INFORMATION: Clone ID: MRT4577_107790C.1.pep
US-10-425-115-193207

Query Match 10.5%; Score 81; DB 6; Length 147;
Best Local Similarity 24.6%; Pred. No. 6.6;
Matches 34; Conservative 18; Mismatches 68; Indels 18; Gaps 4;

QY 7 AAFPAIVVSGSALAGVVPQWGGGNGHNGSSGPDSTLSIYQYGSANAALALQSDARKS 66
DB 27 AQIPASAITGAVAATNPQGGSGTGGVVGAGGPDAPLELYMHD----ILGGSSPTARP 82

QY 67 ETTITQSGYNGADYDQLVTRVVTHEMAHAFRNATIDQWNAKNSDITVQYGGNNAALV 126
DB 83 ITGLGNIYNGQVFPFARIG-----FSAPRNGVAIFNANGQ-----VPTYNGTGIPL 130

QY 127 NCTASDSSVMVQVGFGN 144
DB 131 DTGLSRAGEL--QPTGN 146

RESULT 10
US-10-805-394-4894
Sequence 4894, Application US/10805394
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/10/805,394
CURRENT FILING DATE: 2004-03-22
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 4894
LENGTH: 224
TYPE: PRT
ORGANISM: Corynebacterium glutamicum
US-10-805-394-4894

Query Match 10.5%; Score 81; DB 6; Length 224;
Best Local Similarity 24.6%; Pred. No. 11;
Matches 35; Conservative 19; Mismatches 54; Indels 34; Gaps 4;

QY 16 GSALAGVVPQWGGGNGHNGSS-----GPDSTLSIYQYGSANAALALQSDARKSETTIT 71
DB 45 GTAIAGVLTKGATVEHIGSAUSDIATNGDVVILAVPYPAVESIIASHKDALAKTVID 104

QY 72 QSGYNGADYDQLV-----TRVVTHEMAHAFRN--ATI----- 103
DB 105 ITNPLNFETFDLSLVFVGVSSATAETCAQLPTSRVLKAFNTNFAATLATGKVGDIITTVLV 164

QY 104 --DOWNAKNSDITVQYGGNNA 123
DB 165 AGDEDAKNAITDYNAGGLDA 186

RESULT 11
US-09-248-796A-26659
Sequence 26659, Application US/09248796A
GENERAL INFORMATION:
APPLICANT: Keith Weinstock et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.132
CURRENT APPLICATION NUMBER: US/09/248,796A
CURRENT FILING DATE: 1999-02-12
PRIOR APPLICATION NUMBER: US 60/074,725
PRIOR FILING DATE: 1998-02-13
PRIOR APPLICATION NUMBER: US 60/096,409
PRIOR FILING DATE: 1998-08-13
NUMBER OF SEQ ID NOS: 28208
SEQ ID NO 26659
LENGTH: 457
TYPE: PRT
ORGANISM: Candida albicans
US-09-248-796A-26659

Query Match 10.5%; Score 81; DB 5; Length 457;
Best Local Similarity 25.0%; Pred. No. 25;
Matches 40; Conservative 19; Mismatches 69; Indels 32; Gaps 6;

Qy 15 SGSALAGVVPWGGGNGHGGNS-----GPDSTLSIYQYGSANAALALQSDARKSET 68
Db 289 SSEFLASLINGLGGGGGNGSGNTSNYKNHSTTSTTSKYFNSSTATKLSSSKSIYVNS 348
Qy 69 TITOSGY-----GNGADYDQ-----VTRVVTHEMAFAFRNATIDOWNAKNSD 112
Db 349 TTSRSSISVSSSSDTGGGGANLFGLLNSVAASRTLAESTLSTGTTTSDSANGSTKD 408
Qy 113 ITVQYGGNNAALVNQTASDSSVMVRQVGFQNNAT-ANQY 151
Db 409 YS--SYSGTITSFSTTCSLS-----GDGNKLGNGKY 439

RESULT 12

PCT-US04-21492-458
; Sequence 458, Application PC/TUS0421492
; GENERAL INFORMATION:
; APPLICANT: Diversa Corporation
; APPLICANT: Steer, Brian
; APPLICANT: Callen, Walter
; APPLICANT: Healey, Shaun
; APPLICANT: Pulliam, Derrick
; TITLE OF INVENTION: GLUCANASES, NUCLEIC ACIDS ENCODING THEM AND METHODS FOR MAKING AN
; TITLE OF INVENTION: THEM
; FILE REFERENCE: 564462009540
; CURRENT APPLICATION NUMBER: PCT/US04/21492
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: 60/484,725
; PRIOR FILING DATE: 2003-07-02
; NUMBER OF SEQ ID NOS: 518
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 458
; LENGTH: 568
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from environmental sample
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)...(39)
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (46)...(307)
; OTHER INFORMATION: Cellulase (glycosyl hydrolase family 5)
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (470)...(567)
; OTHER INFORMATION: Cellulose binding domain
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (385)...(413)
; OTHER INFORMATION: Cellulose or protein binding domain

PCT-US04-21492-458

Query Match 10.5%; Score 81; DB 1; Length 568;
Best Local Similarity 20.7%; Pred. No. 32;
Matches 35; Conservative 22; Mismatches 52; Indels 60; Gaps 7;
Qy 26 WGGGNGHN-----GGNNSGPDSTLSIYQYGSANAALALQSDARKSET 68
Db 403 WGEGNQSCVAVSTCSSQPAPYGVVGGSSSGMSSSVAPSSSSSTVSLPASSSSSS 462
Qy 69 TITOSGYNGADYDQVTRVVTHEMAFA-----RNNATIDOWNAK----- 109
Db 463 SSTASGDGN-CEY-----IVSNEWNTGFTCAIRITNEGSAINGMNVNSYSYDGTSVTS 515
Qy 110 -----NSDITVQYGGNNAALVNQTASDSSVMVRQVGF-GNNATAN 149
Db 516 AWMATVSGNNPYASNLGNWGTIQPGQTV-----EFGQGNKGSN 556

RESULT 13

US-10-170-205E-16659
; Sequence 16659, Application US/10170205E
; GENERAL INFORMATION:
; APPLICANT: ADAMS, Mark
; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN
; TITLE OF INVENTION: CAPTURE AGENTS, AND USES THEREOF
; FILE REFERENCE: CLO01381
; CURRENT APPLICATION NUMBER: US/10/170,205E
; CURRENT FILING DATE: 2002-06-13
; NUMBER OF SEQ ID NOS: 40312
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 16659
; LENGTH: 1627
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-170-205E-16659

Query Match 10.3%; Score 80; DB 6; Length 1627;
Best Local Similarity 26.3%; Pred. No. 1.4e+02;
Matches 41; Conservative 16; Mismatches 51; Indels 48; Gaps 9;
Qy 10 AAIIVSGSALAGVVPQ-----WGGGNGHN-----GGNNSGPDSTLSI 47
Db 621 AAAAXSGHAWSGAANQEDKSPTWGPPPKPKSCHWGDGQSRNPASAGGDWADSSSVLGH 680
Qy 48 YOYGSANAALALQSDARKS-----ETTTT-QSGYNGADYDQVTRVVTHEMAFAFRN-N 100
Db 681 LGDGKKNGS-GWDADSNRSGSGWNDDTRSGNSGWNS-----TNTKANPQTNWG 728
Qy 101 ATI-----DQWNAKNSDITVQYGGNNAALVNQTAS 131
Db 729 ETLKPGPQQNWASKPDNNVSNWGG--AASVKQTGT 762

RESULT 14

US-10-425-115-312468
; Sequence 312468, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 312468
; LENGTH: 295
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_48027C.1.pap
US-10-425-115-312468

Query Match 10.2%; Score 79; DB 6; Length 295;
Best Local Similarity 25.4%; Pred. No. 23;
Matches 32; Conservative 13; Mismatches 43; Indels 38; Gaps 5;
Qy 26 WGGGNGHNGGNSGPDSTLSIYQYGSANAALALQSDARKSETTITQSCGYNGADYDQIV 85
Db 138 YGGGGYSGGGYSSG-GYAANGYGVGSGNYSNAGGGYSGS---DGYNGA----- 186
Qy 86 TRVVTHEMAFAFRNATIDQWNAKNSDITVQYGGNNAALVNQTASDSSVMVRQVGFQNN 145
Db 187 -----ASGGYANNLS-SGYNNGRYNTIGSSDGTG-----GYN 219
Qy 146 ATANQY 151
Db 220 SYPNPY 225

```

RESULT 15
PCT-US04-21492-402
; Sequence 402, Application PC/TUS0421492
; GENERAL INFORMATION:
; APPLICANT: Diversa Corporation
; APPLICANT: Steer, Brian
; APPLICANT: Callen, Walter
; APPLICANT: Healey, Shaun
; APPLICANT: Pulliam, Derrick
; TITLE OF INVENTION: GLUCANASES, NUCLEIC ACIDS ENCODING THEM AND METHODS FOR MAKING AN
; TITLE OF INVENTION: THEM
; FILE REFERENCE: 564462009540
; CURRENT APPLICATION NUMBER: PCT/US04/21492
; CURRENT FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: 60/484,725
; PRIOR FILING DATE: 2003-07-02
; NUMBER OF SEQ ID NOS: 518
; SOFTWARE: Patent In version 3.1
; SEQ ID NO 402
; LENGTH: 1010
; TYPE: PRT
; ORGANISM: Bacteria
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)...(30)
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (393)...(428)
; OTHER INFORMATION: Carbohydrate binding domain
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (39)....(300)
; OTHER INFORMATION: Cellulase (glycosyl hydrolase family 5)
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (610)...(959)
; OTHER INFORMATION: Glycosyl hydrolases family 6
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (493)...(521)
; OTHER INFORMATION: Cellulose or protein binding domain
PCT-US04-21492-402

```

	Query Match	10.2%; Score 79; DB 1; Length 1010;
	Best Local Similarity	20.8%; Pred. No. 96;
	Matches	32; Conservative 27; Mismatches 45; Indels 50; Gaps 7;
QY	12 IVVGSALAGVVPWG---CGGNHNGGNSGGPDSTLSIYQYSGSANAALCSDARKSET	68
DB	316 LTAGTIVKNIIVSWGTTCINGSSSSSSSSSSSS-----SSSSSSSSSSSSSSSSSS	367
QY	69 TITQSYGYN-----GADY-----DQLVTRVVTMAHAFFRNATTIDQW	106
DB	368 SSGSTGGNCAGVVYPNWTARDWSGAYNHANAGDMV-----YQNSLYRANW	416
QY	107 ---NAKUSD---ITVGYQGNNAAVLNQATSDSS	134
DB	417 YTNVSPGSDASWTSLSGACGGNGSTTTSSSSSSSS	450

Search completed: August 2, 2004, 15:29:55
Job time : 17.8 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.
OM protein - protein search, using sw model
Run on: August 2, 2004, 14:39:53 ; Search time 9.4 seconds
1545.204 Million cell updates/sec
(without alignments)

Title: US-09-543-407-28
Perfect score: 775
Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVFGNNATANQY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78.*
1: PIR1.*
2: PIR2.*
3: PIR3.*
4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB ID	Description
1	683	88.1	151	2	JC6039	fibrin protein ag
2	683	88.1	151	2	A10635	major curlin chain
3	509	65.7	151	2	S70788	curlin protein csg
4	487.5	62.9	152	2	D90806	curlin major subun
5	487.5	62.9	152	2	H85665	hypothetical prote
6	107	13.8	2174	2	E95985	hypothetical Glyci
7	97	12.5	1748	2	S42136	chjB protein - fet
8	94.5	12.2	151	2	JC6040	fibrin protein ag
9	94	12.1	645	2	F70825	probable PPE prote
10	91.5	11.8	602	1	PL0221	leishmanolysin (EC
11	91	11.7	573	2	C86266	Fif19.21 protein -
12	91	11.7	615	2	E70663	probable PPE prote
13	89.5	11.5	151	2	S70787	curlin nucleator p
14	89.5	11.5	151	2	C90806	minor curlin subun
15	89.5	11.5	151	2	G85665	curlin minor chain
16	89.5	11.5	1322	2	S70753	ice nucleation pro
17	89.5	11.5	1567	2	S11672	ice nucleation pro
18	89	11.5	354	2	B70663	probable PPE prote
19	88	11.4	407	2	T21956	hypothetical prote
20	87.5	11.3	447	2	G84687	probable disease r
21	87.5	11.3	1034	2	JC2143	ice nucleation act
22	87.5	11.3	1258	2	JQ0188	ice nucleation pro
23	87	11.2	590	2	E70946	probable PPE prote
24	86.5	11.2	145	2	AD3143	conserved hypothet
25	86.5	11.2	145	2	H98144	hypothetical prote
26	86.5	11.2	151	2	AH0635	nucleation compone
27	86	11.1	590	1	A45621	leishmanolysin (EC
28	85.5	11.0	1053	2	B70987	probable PPE prote
29	85.5	11.0	2204	2	A70524	probable PPE prote

conserved hypothet
hypothetical prote
leishmanolysin (EC
probable lipoprote
probable lipoprote
hypothetical prote
leishmanolysin (EC
leishmanolysin (EC
hemolysin A - Ser
probable PPE prote
probable PPE prote
hypothetical cell
hypothetical prote
related to C2H2 z1
protein F39D8.1b (i
hypothetical prote

ALIGNMENTS

RESULT 1

JC6039
fibrin protein agfa precursor - Salmonella enteritidis
C;Species: Salmonella enteritidis
C;Date: 31-Dec-1996 #sequence_revision 31-Dec-1996 #text_change 08-Oct-1999
C;Accession: JC6039; PC6015; A44898
R;Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Baner, P.A.; Kay, W.W.
J. Bacteriol. 178, 662-667, 1996
A;Title: Salmonella enteritidis agfBAC operon encoding thin, aggregative fimbriae.
A;Reference number: JC6039; MUID:96146512; PMID:9550497
A;Accession: JC6039
A;Molecule type: DNA
A;Residues: 1-151 <COL>
A;Cross-references: GB:U43280; NID:gl184712; PIDN:AA043599.1; PID:gl184714
A;Accession: PC6015
A;Molecule type: protein
A;Residues: 21-52 <CO2>
A;Experimental source: strain 27655-3b
A;Note: the authors translated the codon ACG for residue 44 as Ile
R;Collinson, S.K.; Emody, L.; Muller, K.H.; Trust, T.J.; Kay, W.W.
J. Bacteriol. 173, 4773-4781, 1991
A;Title: Purification and characterization of thin, aggregative fimbriae from Salmonell
A;Reference number: A44898; MUID:91310586; PMID:1677357
A;Contents: 27655
A;Accession: A44898
A;Status: preliminary
A;Molecule type: protein
A;Residues: 21-33 <CO3>
A;Note: sequence extracted from NCBI backbone (NCBIP:45936)
C;Genetics:
A;Gene: agfa
C;Function:
A;Description: major component of thin aggregative fimbriae
A;Note: fimbriae bind to fibronectin, plasminogen, tissue plasminogen activator
C;Keywords: fimbria
F;1-20/Domain: signal sequence #status predicted <SIG>
F;21-151/Product: fibrin protein agfa #status experimental <VAT>

Query Match 88.1%; Score 683; DB 2; Length 151;
Best Local Similarity 90.1%; Pred. No. 1.4e-51;
Matches 136; Conservative 1; Mismatches 14; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNSGSPDSTLSIYOYGSANAALAQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNSGSPDSTLSIYOYGSANAALAQ 60
QY 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAFAFRNNATIDQWNAKNSDITVQYGG 120
Db 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAFAFRNNATIDQWNAKNSDITVQYGG 120
QY 121 NNALVNTQASDSSVMVRQVFGNNATANQY 151

Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 2

AI0635
Major curlin chain precursor [imported] - Salmonella enterica subsp. enterica serovar Typhimurium
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AI0635
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davies, P.; Davies, R.M.; Dowd, L.; Farrar, M.; Mouton, S.; O'Garra, P.
Nature 413, 848-852, 2001
A:Authors: Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davies, P.; Davies, R.M.; Dowd, L.; Farrar, M.; Mouton, S.; O'Garra, P.
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhimurium
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AI0635
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-151 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD08268.1; PID:g16502315; GSPDB:GN00176
C:Genetics:
A:Gene: STY1181

Query Match 88.1%; Score 683; DB 2; Length 151;
Best Local Similarity 90.1%; Pred. No. 1.4e-51;
Matches 136; Conservative 1; Mismatches 14; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDSTLSIYQYGSANAALALQ 60
|||||
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDSTLSIYQYGSANAALALQ 60
QY 61 SPARKSETTITQSGYNGADYDQLVTRVVTTHMAHAFNNATIDOWNAKNSDITVQYGG 120
|||||
Db 61 SPARKSETTITQSGYNGADYDQLVTRVVTTHMAHAFNNATIDOWNAKNSDITVQYGG 120
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
|||||
Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 3

S70788
curlin protein csgA precursor - Escherichia coli (strain K-12)
N:Alternate names: csgA protein; major curlin protein
C:Species: Escherichia coli
C:Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 01-Mar-2002
C:Accession: S70788; G64846; S31202; S34550; S34559
R:Hammar, M.; Arngvist, A.; Bian, Z.; Olsen, A.; Normark, S.
Mol. Microbiol. 18, 661-670, 1995
A:Title: Expression of two csg operons is required for production of fibronectin- and CsgA
A:Reference number: S70783; MUID:96414468; PMID:8817489
A:Accession: S70788
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-151 <HAM>
A:Cross-references: EMBL:X90754; NID:g1147558; PIDN:CAA62282.1; PID:g1147564
A:Experimental source: strain K12, substrain W3110
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, J.; Rose, D.J.; Mau, B.; Shaoy, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: G64846
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-151 <BLAT>
A:Cross-references: GB:AE000205; GB:U00096; NID:g1797265; PIDN:AACT4126.1; PID:g1797279;
A:Experimental source: strain K-12, substrain MGL655
R:Olsen, A.; Arngvist, A.; Hammar, M.; Sukupolvi, S.; Normark, S.
Mol. Microbiol. 7, 523-536, 1993

A:Title: The RpoS sigma factor relieves H-NS-mediated transcriptional repression of csgA
A:Reference number: S31202; MUID:93211294; PMID:8459772

A:Accession: S31202
A:Molecule type: DNA
A:Residues: 1-6, 'V', 8-151 <OLSI>
A:Cross-references: EMBL:L04979
A:Accession: S34560
A:Molecule type: protein
A:Residues: 21-42; 44-50 <OLS2>
R:Olsen, A.N.; Arngvist, A.M.
submitted to the EMBL Data Library, October 1992
A:Reference number: S34559
A:Accession: S34559
A:Molecule type: DNA
A:Residues: 1-133, 'PQRSGMW' <OLS3>
A:Cross-references: EMBL:L04979; NID:G290424; PIDN:AAA23616.1; PID:g290425
A:Experimental source: strain K-12, substrain W3110
C:Genetics:
A:Gene: csgA
A:Map position: 23.15
C:Function:
A:Description: major component of wild-type curli; interaction between CsgA and CsgB tri
A:Note: curli are thin, coiled fibers expressed on the surface of Escherichia coli that
and H-kininogen; in the absence of CsgA, CsgB can self-assemble into polymers
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-151/Product: curlin #status experimental <MAT>

Query Match 65.7%; Score 509; DB 2; Length 151;
Best Local Similarity 68.2%; Pred. No. 1.1e-36;
Matches 103; Conservative 15; Mismatches 33; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDSTLSIYQYGSANAALALQ 60
|||||
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDSTLSIYQYGSANAALALQ 60
QY 61 SPARKSETTITQSGYNGADYDQLVTRVVTTHMAHAFNNATIDOWNAKNSDITVQYGG 120
|||||
Db 61 SPARKSETTITQSGYNGADYDQLVTRVVTTHMAHAFNNATIDOWNAKNSDITVQYGG 120
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
|||||
Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 4

D90806
curlin major subunit CsgA [imported] - Escherichia coli (strain O157:H7, substrain RIMD
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: D90806
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: D90806
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-152 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA34843.1; PID:g13360880; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetics:
A:Gene: ECS1420

Query Match 62.9%; Score 487.5; DB 2; Length 152;
Best Local Similarity 66.4%; Pred. No. 7.6e-35;
Matches 101; Conservative 16; Mismatches 34; Indels 1; Gaps 1;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDSTLSIYQYGSANAALAL 59
|||||
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDSTLSIYQYGSANAALAL 60
QY 60 QSDARKSETTITQSGYNGADYDQLVTRVVTTHMAHAFNNATIDOWNAKNSDITVQYGG 119

[illegible]

A;Molecule type: DNA
A;Residues: 1-1748 <TAV>
A;Cross-references: EMBL:L03710; NID:G161751; PID:G161752
R;Taylor, F.M.; Martindale, D.W.
Nucleic Acids Res. 21, 4610-4614, 1993
A;Title: Retroviral-type zinc fingers and glycine-rich repeats in a protein encoded by the *src* gene
A;Reference number: S42135; MUID:94051569; PMID:8233798
A;Accession: S42135
A;Molecule type: DNA
A;Residues: 1164-1174; 1179-1198; 1233-1252; 1285-1293; 1297-1309; 1316-1326; 1331-1338
A;Cross-references: EMBL:L03710
R;Martindale, D.W.; Taylor, F.M.
Nucleic Acids Res. 16, 2189-2201, 1988
A;Title: Multiple introns in a conjugation-specific gene from *Tetrahymena thermophila*
A;Reference number: S03650; MUID:88199811; PMID:3357771

[illegible]

QY 142 FGNNTAN 149
 Db 1723 WGSNNQAS 1730

RESULT 8
 JC5040
 fimbria protein agfB precursor - Salmonella enteritidis
 C:Species: Salmonella enteritidis
 C>Date: 31-Dec-1996 #sequence_revision 31-Dec-1996 #text_change 08-Oct-1999
 C:Accession: JC6040
 R:Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Baner, P.A.; Kay, W.W.
 J. Bacteriol. 178, 662-667, 1996
 A:Title: Salmonella enteritidis agfBAC operon encoding thin, aggregative fimbriae.
 A:Reference number: JC6039; MUID:96146512; PMID:8550497
 A:Molecule type: DNA
 A:Residues: 1-151 <COL>
 A:Cross-references: GB:U43280; NID:g1184712; PIDN:AAC43598.1; PID:g1184713
 A:Experimental source: strain 276755-3b
 C:Genetics:
 A:Gene: agfB
 C:Function:
 A:Description: minor component of thin aggregative fimbriae
 A:Note: fimbriae bind to fibronectin, plasminogen, tissue plasminogen activator
 C:Keywords: fimbria
 F:1-21/Domain: signal sequence #status predicted <SIG>
 F:22-151/Product: fimbria protein agfB #status predicted <MAT>

Query Match 12.2%; Score 94.5; DB 2; Length 151;
 Best Local Similarity 26.9%; Pred. No. 0.34;
 Matches 35; Conservative 17; Mismatches 41; Indels 37; Gaps 4;

QY 10 RAIIVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALQSDARKSETT 69
 Db 58 ARVQEGSKLSVLSQ--EGGNRAKVQDAGNYFAYIEQTGNAN-----DAS 103

QY 70 ITQSGYNGADYDQDLVTRVVTHEMAFRNNATIDOWNAKNSDITVGYGGNNAALVNOT 129
 Db 104 ISQSAYGNSA-----AIQKSGNKANIT--QYGTQRTAVVQX 140

QY 130 ASDSSVMVQ 139
 Db 141 QSHMAIRVTQ 150

RESULT 9
 F70825
 Probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
 C:Species: Mycobacterium tuberculosis
 C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 15-Sep-2003
 C:Accession: F70825
 R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
 A:Authors: Squares, R.; Sulston, J.B.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A:Reference number: A70500; MUID:98295987; PMID:9634230
 A:Accession: F70825
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-645 <COL>
 A:Cross-references: GB:AL021958; GB:AL123456; NID:g3261536; PIDN:CAAL17522.1; PID:e125329
 A:Experimental source: strain H37RV
 C:Genetics:
 A:Gene: PPE

Query Match 12.1%; Score 94; DB 2; Length 645;
 Best Local Similarity 24.3%; Pred. No. 2;
 Matches 34; Conservative 14; Mismatches 42; Indels 50; Gaps 7;

QY 11 AIIVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALQSDARKSETT 69
 Db 294 AVPTPGNGNVGI-----GNGGNGFGGNT-----GNANIGLGNVGDG----- 331

QY 70 ITQSGYNGADYDQDLVTRVVTHEMAFRNNATIDOWNAKNSDITVGYGGNNAALVNOT 129
 Db 332 --NVGFGNSGSYN-----FGFGNTG-----NNNIGLGLTSQNGIIFGGGLN 369

QY 130 ASDSSVMVQVQFGNNATAN 149
 Db 370 SGSGN-----IGFGNSGTGN 384

RESULT 10
 PL0221
 Leishmanolysin (EC 3.4.24.36) precursor [validated] - Leishmania major
 N:Alternate names: promastigote surface proteinase; surface endopeptidase glycoprotein
 C:Species: Leishmania major
 C>Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 15-Sep-2000
 C:Accession: PL0221; A27598; A60648
 R:Button, L.L.; McMaster, W.R.
 J. Exp. Med. 171, 589, 1990
 A:Reference number: PL0221
 A:Contents: erratum
 A:Molecule type: DNA
 A:Residues: 1-602 <BUT>
 A:Cross-references: GB:Y00647; NID:g9554; PIDN:CAA68673.1; PID:g9555
 A:Note: this is a revision to the sequence from reference A27598
 R:Button, L.L.; McMaster, W.R.
 J. Exp. Med. 167, 724-729, 1988
 A:Title: Molecular cloning of the major surface antigen of Leishmania.
 A:Reference number: A27598; MUID:88154764; PMID:3346625
 A:Accession: A27598
 A:Status: significant sequence differences
 A:Molecule type: DNA
 R:Bouvier, J.; Bordier, C.; Vogel, H.; Reichelt, R.; Etges, R.
 Mol. Biochem. Parasitol. 37, 235-246, 1989
 A:Title: Characterization of the promastigote surface protease of Leishmania as a membra
 A:Reference number: A60648; MUID:90114330; PMID:2608099
 A:Accession: A60648
 A:Molecule type: protein
 A:Residues: 101, 'E', 103-118, 'SV', 121-123 <BOU>
 A:Experimental source: strain LEM513
 R:Schlagerhauf, E.; Etges, R.; Metcalf, P.
 submitted to the Brookhaven Protein Data Bank, March 1997
 A:Reference number: A68135; PDB:1LMV
 A:Contents: annotation; X-ray crystallography, 1.86 angstroms, residues 100-407, 412-498,
 A:Note: strain LRC-L119
 C:Complex: homodimer
 C:Function:
 A:Description: catalyzes the hydrolysis of peptide bonds between two hydrophobic residues
 A:Note: the activated form can activate the proenzyme form
 C:Superfamily: leishmanolysin
 C:Keywords: blocked carboxyl end; cell adhesion; glycoprotein; homodimer; hydrolase; lip
 F:1-33/Domain: signal sequence #status predicted <SIG>
 F:40-100/Domain: activation peptide #status predicted <ATP>
 F:101-577/Product: leishmanolysin #status experimental <MAT>
 F:578-602/Domain: carboxyl-terminal propeptide #status predicted <CTP>
 F:48, 264, 268, 334/Binding site: zinc, catalytic (Cys, His, His) (inhibited) #status
 F:100-101/Cleavage site: Val-Val (autolytic) #status experimental
 F:125-142, 191-230, 314-386, 393-455, 406-425, 415-499, 466-510, 515-565, 535-558/Disulfide bond
 F:264, 268, 334/Binding site: zinc, catalytic (His) (active) #status experimental
 F:265/Active site: Glu #status predicted
 F:300, 407/Binding site: carboxylate (Asn) (covalent) #status experimental
 F:577/Modified site: GPI-anchor ethanolamine amidated carboxyl end (Asn) (in mature form)

Query Match 11.8%; Score 91.5; DB 1; Length 602;
 Best Local Similarity 41.7%; Pred. No. 3;
 Matches 25; Conservative 2; Mismatches 12; Indels 21; Gaps 2;

QY 81 YDQLVTRVVTHEMAFR-----RNNATIDOWNAKNSDITV-----QQYG 119
 |||||

Db 254 YDQLVTRVTHEMAHALGFGSPFFEDARIIVANVPNVRGNFDVPVINSSTAVAKAREQY 313

RESULT 11

C86266

F3P19_21 protein - Arabidopsis thaliana

C/Species: Arabidopsis thaliana (mouse-ear cress)

C/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Nov-2001

C/Accession: C86266

R/Theologitis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A/Authors: Salzborg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A/Reference number: A86141; MUID:21016719; PMID:11130712

A/Accession: C86266

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-573 <STO>

A/Cross-references: GB:AE005172; NID:g485042; PIDN:AAD31072.1; GSPDB:GN00141

C/Genetics:

A/Map position: 1

Query Match 11.7%; Score 91; DB 2; Length 573;

Best Local Similarity 26.9%; Pred.No. 3.2;

Matches 36; Conservative 19; Mismatches 49; Indels 30; Gaps 7;

Qy 13 VWSGSLAGVVPQWGGGNGNGSGGPDSTLSIYQGSANAALQSDARKSETTITQ 72

Db 84 IVSGGTVEG---KYRNDGGHNG---ISGPTRSVYQASFGAKGLNIIQSNK--IAQ 135

Qy 73 SGYNGAGADYDQLVTRVTHEMAHAFRNA-TIDQWAKNS-----DITVGYQYGN 122

Db 136 QG-----STTWLNN--HGFGNAVNPMPVHNSYGAPPQGAQQTIPVQMSVNP 183

Qy 123 AALVNTQATSDSSVM 136

Db 184 NVMNKSPQTQSFVV 197

RESULT 12

E70663

Probable PPE protein - Mycobacterium tuberculosis (strain H37RV)

C/Species: Mycobacterium tuberculosis

C/Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 15-Sep-2003

C/Accession: E70663

R/Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, K.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A/Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A/Reference number: A70500; MUID:98295987; PMID:9634230

A/Accession: E70663

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-615 <COL>

A/Cross-references: GB:283860; NID:g3261691; PIDN:CAB06165.1; PID:e290763;

A/Experimental source: strain H37RV

C/Genetics:

A/Gene: PPE

Query Match 11.7%; Score 91; DB 2; Length 615;

Best Local Similarity 22.5%; Pred.No. 3.4;

Matches 31; Conservative 14; Mismatches 47; Indels 46; Gaps 5;

Qy 15 SGSALAGVVPQWGGGNGNGSGGPDSTLSIYQGSANAALQSDARKSETTITQSG 74

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:36:12 ; Search time 5.3 Seconds
(without alignments)

1483.508 Million cell updates/sec

Title: US-09-543-407-28

Perfect score: 775

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVGFNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	683	88.1	151	1 CSGA_SALTY	P55225 salmonella
2	509	65.7	151	1 CSGA_ECOLI	P28307 escherichia
3	487.5	62.9	152	1 CSGA_ECOS7	Q93U24 escherichia
4	94.5	12.2	151	1 CSGB_SALTY	P55226 salmonella
5	91.5	11.8	602	1 GP63_LEIMA	P08148 leishmania
6	89.5	11.5	151	1 CSGB_ECOLI	P39828 escherichia
7	89.5	11.5	1322	1 ICBA_PANAN	P20469 pantoea ana
8	89.5	11.5	1567	1 ICEN_XANCT	P18127 xanthomonas
9	87.5	11.3	1034	1 ICEN_PANAN	Q47879 pantoea ana
10	87.5	11.3	1258	1 ICEN_ERWHE	P16239 erwinia her
11	86.5	11.2	151	1 CSGB_SALTY	Q827m3 salmonella
12	86	11.1	590	1 GP63_LEIDO	P23223 leishmania
13	84.5	10.9	646	1 GP63_LEIME	P43150 leishmania
14	84	10.8	599	1 GP63_LEICH	P15706 leishmania
15	84	10.8	1608	1 HLVA_SERMA	P15320 serratia ma
16	83.5	10.8	856	1 AE7_AZOVI	Q92f69 azotobacter
17	83	10.7	1656	1 CMPE_RICVA	O08653 r outer mem
18	81.5	10.5	392	1 HMEL_HUMAN	Q05925 homo sapien
19	81.5	10.5	697	1 SIM_DROME	P15709 drosophila
20	81.5	10.5	1028	1 CVO_DRCME	P51521 drosophila
21	80	10.3	331	1 CMB2_NEIMB	P30688 neisseria m
22	80	10.3	760	1 YBIL_ECOLI	P75780 escherichia
23	79.5	10.3	475	1 ABPB_STRCO	O51161 streptomyce
24	79.5	10.3	548	1 CEAK_ECOLI	Q47502 escherichia
25	79.5	10.3	1210	1 ICEN_PSEFL	P09815 pseudomonas
26	79	10.2	678	1 YF48_MYCTU	Q10778 mycobacteri
27	79	10.2	823	1 NSP1_YEAST	P14907 saccharomyc
28	78.5	10.1	232	1 DNH3_PEA	P28641 pisum sativ
29	78.5	10.1	641	1 IMD_ABTGO	Q44052 athrobacte
30	77.5	10.0	1063	1 SPF5_YEAST	P27692 saccharomyc
31	77.5	10.0	1148	1 ICXK_PSEBX	O30611 pseudomonas
32	77.5	10.0	1185	1 MAPX_DROME	P23226 drosophila
33	77.5	10.0	1196	1 ICEV_PSEBX	O33479 pseudomonas

ALIGNMENTS

RESULT 1
CSGA_SALTY
ID CSGA_SALTY STANDARD; PRT; 151 AA.
AC P55225;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Major curlin subunit precursor (Fimbrin SEF17).
GN CSGA OR AGFA OR STM1144 OR STY1181 OR T1776.
OS Salmonella typhimurium,
OS Salmonella typhi, and
OS Salmonella enteritidis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602, 601, 592;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=SR-11;
RX MEDLINE=98117058; PubMed=9457880;
RA Romling U., Bian Z., Hammar M., Sierralta W.D., Normark S.;
RT "Curli fibers are highly conserved between Salmonella typhimurium and
Escherichia coli with respect to operon structure and regulation.";
RL J. Bacteriol. 180:722-731(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman P.,
Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
LT2.";
RL Nature 413:852-856(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
Whitehead S., Barrall B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
enterica serovar Typhi CT18";
RL Nature 413:848-852(2001).
RN [4]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=22551367; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;

34 77 9.9 204 1 CORA_MEDSA
35 77 9.9 376 1 P33_MYCPE
36 77 9.9 439 1 ACMA_LACLA
37 76.5 9.9 487 1 Y442_MYCTU
38 76.5 9.9 881 1 YFCU_ECOLI
39 76 9.8 147 1 HFAA_CAUCR
40 76 9.8 362 1 P35_MYCPE
41 76 9.8 542 1 SCWE_YEAST
42 76 9.8 1571 1 C3G_DROME
43 75.5 9.7 163 1 HCY_NAIPH
44 75.5 9.7 394 1 OMSI_SALTY
45 75.5 9.7 424 1 COAA_BFPD
O07202 medicago sa
Q50368 mycoplasma
Q9c1t4 lactococcus
P77196 escherichia
P77342 caulobacter
Q50367 mycoplasma
P53189 saccharomyc
O77086 drosophila
P39442 natronomona
Q56110 salmonella
P03661 bacteriophag

```

RT "Comparative Genomics of Salmonella enterica serovar Typhi strains Ty2
RL and CT18."
RL J. Bacteriol. 185:2330-2337(2003).
RN [5]
RP SEQUENCE FROM N.A.
RC SPECIES=S.enteritidis; STRAIN=27655-3B;
RX MEDLINE=96146512; PubMed=8550497;
RA Collinson S.K., Clouthier S.C., Doran J.L., Bansen P.A., Kay W.W.;
RT "Salmonella enteritidis agfBAC operon encoding thin, aggregative
fimbriae."
RT J. Bacteriol. 178:662-667(1996).
RN [6]
RP SEQUENCE OF 21-151 FROM N.A.
RC SPECIES=S.enteritidis; STRAIN=27655-3B;
RX MEDLINE=94013373; PubMed=8104955;
RA Doran J.L., Collinson S.K., Burian J., Santos G., Todd E.C.D.,
Munro C.K., Kay C.M., Bansen P.A., Peterkin P.I., Kay W.W.;
RT "DNA-based diagnostic tests for Salmonella species targeting agfA,
the structural gene for thin, aggregative fimbriae."
RT J. Clin. Microbiol. 31:2263-2273(1993).
RN [7]
RP SEQUENCE OF 21-33.
RC SPECIES=S.enteritidis; STRAIN=27655-3B;
RX MEDLINE=91310586; PubMed=1677357;
RA Collinson S.K., Emeody L., Mueller K.-M., Trust T.J., Kay W.W.;
RT "Purification and characterization of thin, aggregative fimbriae from
Salmonella enteritidis."
RL J. Bacteriol. 173:4773-4781(1991).
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN.
CC -!- SIMILARITY: BELONGS TO THE CSGB/CSGB FAMILY.
CC
CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AJ002301; CAA05317.1; -
DR EMBL; AF008749; AAL20074.1; -
DR EMBL; AL627269; CAD08268.1; -
DR EMBL; A2016840; AAC069399.1; -
DR EMBL; U43280; AAC43599.1; -
DR FIC; JC6039; JC6039.
DR StyGene; SGI0608; csGA.
KW Fimbria; Signal; Complete proteome.
FT SIGNAL 1 20
FT CHAIN 21 151
FT CONFLICT 134 151
FT REF. 6)
FT SEQUENCE 151 AA; 15305 MW; 57DAC0D16B621359 CRC64;
SQ
Query Match 88.1%; Score 683; DB 1; Length 151;
Best Local Similarity 90.1%; Pred. No. 1.3e-51;
Matches 136; Conservative 1; Mismatches 14; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVSGSALAGVVPQWGGGNGHNGGSGPSTLSIIYQGSANAALQ 60
DB 1 MKLLKVAFAAIVSGSALAGVVPQWGGGNGHNGGSGPSTLSIIYQGSANAALQ 60
QY 61 SPARKSETTITSGYXGNGADYDQLVTRVVVTHEMAHAFRNATTIDWNNAKNSDITVQYGG 120
DB 61 SPARKSETTITSGYXGNGADYDQLVTRVVVTHEMAHAFRNATTIDWNNAKNSDITVQYGG 120
QY 121 NNAALVNQTSADSSVMVROVFGNNATANOY 151
DB 121 NNAALVNQTSADSSVMVROVFGNNATANOY 151

```

```

RESULT 2
CSGA_ECOLI STANDARD; PRT; 151 AA.
ID CSGA_ECOLI
AC P28307;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Major curlin subunit precursor.
GN CSGA OR B1042.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / W3110;
RX MEDLINE=93211294; PubMed=8459772;
RA Olsen A., Arngqvist A.;
RT "The RpoS sigma factor relieves H-NS-mediated transcriptional
repression of csGA, the subunit gene of fibronectin-binding curli in
Escherichia coli."
RT Escherichia coli."
RL Mol. Microbiol. 7:523-536(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MC4100;
RX MEDLINE=96414468; PubMed=8817489;
RA Hammar M., Arngqvist A., Bian Z., Olsen A., Normark S.;
RT "Expression of two csG operons is required for production of
fibronectin- and Congo red-binding curli polymers in Escherichia coli
K-12."
RT Mol. Microbiol. 18:661-670(1995).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
Sampel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
Yano M., Horiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 12.7-28.0 min region on the linkage map."
RL DNA Res. 3:137-155(1996).
RN [5]
RP SEQUENCE OF 21-40.
RC STRAIN=K12 / YMEI;
RX MEDLINE=93023873; PubMed=1357528;
RA Arngqvist A., Olsen A., Pfeifer J., Russell D.G., Normark S.;
RT "The Crl protein activates cryptic genes for curli formation and
fibronectin binding in Escherichia coli HB101."
RL Mol. Microbiol. 6:2443-2452(1992).
RN [6]
RP SEQUENCE OF 21-31.
RX MEDLINE=91310586; PubMed=1677357;
RA Collinson S.K., Emeody L., Trust T.J., Kay W.W.;
RT "Purification and characterization of thin, aggregative fimbriae from
Salmonella enteritidis."
RL J. Bacteriol. 173:4773-4781(1991).
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN.
CC

```

```

CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L04979; AAA23616.1; -
CC DR EMBL; X90754; CAA62282.1; -
CC DR EMBL; AE000205; AAC74126.1; -
CC DR EMBL; D90741; BAA35832.1; -
CC DR EMBL; D90742; BAA35840.1; -
CC DR PIR; S70788; S70788.
CC DR EcoGene; EG11489; csGA.
CC KW Fimbria; Signal; Complete proteome.
CC FT SIGNAL 1 20
CC FT CHAIN 21 151 MAJOR CURLIN SUBUNIT.
CC FT CONFLICT 7 7 A -> E (IN REF. 1).
CC SQ SEQUENCE 151 AA; 15049 MW; C003470D208D395P CRC64;

Query Match 65.7%; Score 509; DB 1; Length 151;
Best Local Similarity 68.2%; Pred. No. 8.7e-37;
Matches 103; Conservative 15; Mismatches 33; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGHHNGGNSGDPSTLSIYQYGSANAALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGHHNGGNSGDPSTLSIYQYGSANAALQ 60
QY 61 SDARKSETITQSGYNGADYDQLVTRVVTTHMAHAFNNATIDQWNAKNSDITVGOYGG 120
DB 61 TDARNSDLTITQHGCGGADYQGSDDSSIDLTDQFGNSATLDQWNGKSEMTVKQFGG 120

QY 121 NNAALVNQTSADSSVMVQVFGNNATANQY 151
DB 121 GNGAAVDQTSNNSVNVTVQVFGNNATAHQY 151

RESULT 3
CSGA_ECO57 STANDARD; PRT; 152 AA.
AC Q93U24;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Major curlin subunit precursor.
GN CSGA OR Z1676 OR ECS1420.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / ATCC 43895;
RX MEDLINE=21218556; PubMed=11319125;
RA Ulrich G.A., Keen J.E., Elder R.O.;
RT "Mutations in the csGD promoter associated with variations in curli
RT expression in certain strains of Escherichia coli O157:H7.";
RL Appl. Environ. Microbiol. 67:2367-2370(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoculis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).

```

```

RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COLLED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN.
CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF275733; AK53212.1; -
CC DR EMBL; AE005315; AG55788.1; -
CC DR EMBL; AP002554; BAB34843.1; -
CC DR PIR; D90806; D90806.
CC DR PIR; H85665; H85665.
CC KW Fimbria; Signal; Complete proteome.
CC FT SIGNAL 1 20
CC FT CHAIN 21 152 MAJOR CURLIN SUBUNIT.
CC SQ SEQUENCE 152 AA; 15099 MW; EE2D2D94DDE91243 CRC64;

Query Match 62.9%; Score 487.5; DB 1; Length 152;
Best Local Similarity 66.4%; Pred. No. 6e-35;
Matches 101; Conservative 16; Mismatches 34; Indels 1; Gaps 1;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGHHNGGNSGDPSTLSIYQYGSANAALQ 59
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGHHNGGNSGDPSTLSIYQYGSANAALQ 60
QY 60 QSDARKSETITQSGYNGADYDQLVTRVVTTHMAHAFNNATIDQWNAKNSDITVGOYGG 119
DB 61 QADARNSDLTITQHGCGGADYQGSDDSSIDLTDQFGNSATLDQWNGKSHMTVKQFG 120

QY 120 GNGAAVDQTSNNSVNVTVQVFGNNATANQY 151
DB 121 GNGAAVDQTSNNSVNVTVQVFGNNATAHQY 152

RESULT 4
CSGB_SALTY STANDARD; PRT; 151 AA.
AC P55226;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Minor curlin subunit precursor (Fimbrin SEF17 minor subunit).
GN CSGB OR AGFB OR STM1143.
OS Salmonella typhimurium, and
OS Salmonella enteritidis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602, 592;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=SR-11;
RX MEDLINE=98117058; PubMed=9457880;
RA Romling U., Bian Z., Hammar M., Sierralta W.D., Normark S.;
RT "Curli fibers are highly conserved between Salmonella typhimurium and
RT Escherichia coli with respect to operon structure and regulation.";

```



```
FT METAL 264 264 ZINC (CATALYTIC).
FT ACT_SITE 265 265
FT METAL 268 268 ZINC (CATALYTIC).
FT METAL 334 334 ZINC (CATALYTIC).
FT DISULFID 125 142
FT DISULFID 191 230
FT DISULFID 314 386
FT DISULFID 333 455
FT DISULFID 406 425
FT DISULFID 415 489
FT DISULFID 466 510
FT DISULFID 515 565
FT DISULFID 535 558
FT CARBOHYD 300 300
FT CARBOHYD 407 407
FT LIPID 577 577
FT STRAND 101 102
FT STRAND 107 108
FT STRAND 111 114
FT HELIX 116 119
FT TURN 121 122
FT TURN 128 129
FT STRAND 131 133
FT STRAND 139 141
FT HELIX 144 146
FT STRAND 150 158
FT TURN 159 159
FT HELIX 160 169
FT TURN 170 171
FT STRAND 172 174
FT STRAND 177 178
FT STRAND 180 181
FT TURN 189 190
FT HELIX 191 193
FT HELIX 198 202
FT TURN 203 203
FT STRAND 205 206
FT STRAND 210 215
FT TURN 221 222
FT STRAND 226 232
FT TURN 234 235
FT STRAND 238 244
FT HELIX 247 249
FT HELIX 256 269
FT TURN 270 271
FT HELIX 274 279
FT TURN 280 281
FT STRAND 283 286
FT HELIX 289 291
FT STRAND 296 299
FT HELIX 302 312
FT TURN 313 313
FT TURN 315 316
FT STRAND 320 322
FT TURN 328 332
FT STRAND 335 335
FT TURN 337 339
FT TURN 341 342
FT STRAND 343 343
FT TURN 344 345
FT STRAND 353 353
FT HELIX 356 364
FT TURN 365 366
FT STRAND 369 370
FT HELIX 372 374
FT TURN 380 383
FT HELIX 386 390
FT STRAND 394 395
FT TURN 396 397
FT STRAND 398 399
FT TURN 402 404
FT STRAND 413 414
FT TURN 417 418

ZINC (CATALYTIC).
ZINC (CATALYTIC).
ZINC (CATALYTIC).

N-LINKED (GLCNAC. . .) (POTENTIAL).
N-LINKED (GLCNAC. . .) (POTENTIAL).
GPI-anchor amidated asparagine.

FT STRAND 421 425
FT STRAND 428 429
FT HELIX 435 437
FT TURN 443 444
FT STRAND 445 446
FT TURN 450 454
FT STRAND 458 465
FT TURN 466 467
FT HELIX 470 472
FT TURN 475 477
FT HELIX 478 480
FT TURN 485 486
FT STRAND 487 494
FT STRAND 496 496
FT STRAND 506 516
FT TURN 517 520
FT STRAND 521 525
FT TURN 527 528
FT STRAND 533 534
FT TURN 537 538
FT STRAND 540 542
FT HELIX 543 545
FT TURN 546 546
FT STRAND 550 550
FT TURN 552 553
FT STRAND 555 557
FT HELIX 561 565
FT TURN 566 567
FT HELIX 569 572
FT TURN 573 573
SQ SEQUENCE 602 AA; 63953 MW; 982BF3245D87C43E CRC64;

Query Match 11.8%; Score 91.5; DB 1; Length 602;
Best Local Similarity 41.7%; Pred. NO. 1.6;
Matches 25; Conservative 2; Mismatches 12; Indels 21; Gaps 2;

QY 81 YDQLVTRVVTHEMAHAF-----RNNATIDWNAXNSDITV-----GOYG 119
      |||||
Db 254 YDQLVTRVVTHEMAHALGFGPGFFEDARIVANVNVGRGNFDVPVINSSTAVAKAEQYG 313

RESULT 6
CSGB_ECOLI STANDARD; PRT; 151 AA.
AC P39828;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Minor curlin subunit precursor.
GN CSGB OR B1041 OR Z1675 OR ECS1419.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MC4100;
RX MEDLINE=96414468; PubMed=8817489;
RA Hammar M., Arnoqvist A., Bian Z., Olsen A., Normark S.;
RT "Expression of two csg operons is required for production of
  fibronectin- and congo red-binding curli polymers in Escherichia coli
  K-12.";
RT K-12.";
RL Mol. Microbiol. 18:661-670(1995).
RP [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MC1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
```

```

RL Science 277:1453-1474 (1997).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikenoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RA "A 718-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA Res. 3:1137-1155 (1996).
RN [4]
RP SEQUENCE FROM N.A.
RP STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.P., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Foster G., Hackett J., Klink S., Boulin A., Shao Y., Miller L.,
RA Grobbeck E.J., Davis N.W., Lim A., Dimmalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RA "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533 (2001).
RN [5]
RP SEQUENCE FROM N.A.
RP STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Lida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RA "Complete genome sequence of enterohaemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22 (2001).
RN [6]
RP SEQUENCE OF 1-21 FROM N.A.
RP STRAIN=K12;
RX MEDLINE=95157246; PubMed=7854117;
RA Arngqvist A., Olsen A., Normark S.;
RA "Sigma S-dependent growth-phase induction of the esgBA promoter in
Escherichia coli can be achieved in vivo by sigma 70 in the absence
of the nucleoid-associated protein H-NS.";
RL Mol. Microbiol. 13:1021-1032 (1994).
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRINECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
CC CURLIN MONOMERS.
CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X30754; CAA62281.1; -
CC EMBL; AB000205; AAC74125.1; -
CC EMBL; D90741; BAA35831.1; -
CC EMBL; AC005315; AAG55787.1; -
CC EMBL; AP002554; BAB34842.1; -
CC PIR; C90806; C90806.
CC PIR; G85665; G85665.
CC PIR; S70787; S70787.
CC EcoGene; EG12621; csGB.
CC FimBria; Signal; Complete proteome.
CC SIGNAL 1 21 POTENTIAL.
CC CHAIN 22 151 MINOR CURLIN SUBUNIT.

```

```

SQ SEQUENCE 151 AA; 15882 MW; B18D266B964014B8 CRC64;
Query Match 11.5%; Score 89.5; DB 1; Length 151;
Best Local Similarity 28.1%; Pred. No. 0.5; Mismatches 62; Indels 5; Gaps 2;
Matches 32; Conservative 15;
QY 38 SSGPSTLSIYQYGSANAALALQSDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAF 97
Db 21 AAGYDLANSEYNF-----AVNELSKSFNQAAIIGAGTNNNSAQLRQGGSKLLAVVAQSGS 76
QY 98 RNNATIDQWNAKNSDITVGCYQGNNAALVNQTASDSSVMVROVGFNNATANQY 151
Db 77 SNRAKIDQTGDYNL-AYIDQAGSANDASISQAGYNTAMIIQKSGNKANITQY 129
RESULT 7
ICEA_PANAN STANDARD; PRT; 1322 AA.
AC P20469;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ice nucleation protein inaa.
GN INAA.
OS Pantoea ananas (Erwinia uredovora).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pantoea.
OX NCBI_TaxID=553;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9002494; PubMed=25990935;
RA Abe K., Watabe S., Emori Y., Watanabe M., Arai S.;
RA "An ice nucleation active gene of Erwinia ananas. Sequence similarity
to those of Pseudomonas species and regions required for ice
nucleation activity.";
RL FEBS Lett. 258:297-300 (1989).
CC -!- FUNCTION: Ice nucleation proteins enable bacteria to nucleate
CC crystallization in supercooled water. (By similarity).
CC -!- SUBCELLULAR LOCATION: Outer membrane (By similarity).
CC -!- DOMAIN: CONTAINS MANY IMPERFECT REPEATS OF THE CONSENSUS
CC OCTAPEPTIDE A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A
CC REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.
CC -!- SIMILARITY: Belongs to the bacterial ice nucleation protein
CC family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X17316; CAA35194.1; -
CC PIR; S07053; S07053.
CC HSP; P05620; IINA.
CC InterPro; IPR000259; Ice_nucleatn.
CC Pfam; PF00818; Ice_nucleation; 69.
CC PRINTS; PR00327; ICENUCLEATN.
CC PROSITE; PS00314; ICE_NUCLEATION; 49.
CC Ice nucleation; Repeat; Outer membrane.
CC DOMAIN 162 1281 OCTAPEPTIDE PERIODICITY.
CC SEQUENCE 1322 AA; 131094 MW; 89B0EBE24AA837039 CRC64;
Query Match 11.5%; Score 89.5; DB 1; Length 1322;
Best Local Similarity 26.8%; Pred. No. 5.9;
Matches 34; Conservative 18; Mismatches 26; Indels 49; Gaps 7;
QY 34 GGCNSGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGADYDQLVTRVVTHEM 93
Db 933 GSTSTAGPDSL-IAGYGSTQTA-----GYNSILTAGYS----- 966
QY 94 AFAFRNNATIDQWNAKNSDITVGCYG-----GNNAALV-----NQTASDSSVMVROVGF 143

```



```
Db 967 -----TGTGENSEDLTTG-YGSTSTAGYESSLIAGYGSTQTASFXTLM--AGYG 1013
QY 144 NNATANQ 150
Db 1014 SSQTARE 1020

RESULT 8
ICEN_XANCT
ID ICEN_XANCT STANDARD; PRT; 1567 AA.
AC P18127;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ice nucleation protein.
GN INAX.
OS Xanthomonas campestris (pv. translucens).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=343;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=X56S;
RX MEDLINE=91080859; PubMed=2259339;
RA Zhao J., Orser C.S.;
RT "Conserved repetition in the ice nucleation gene inax from
Xanthomonas campestris pv. translucens.";
RL Mol. Genet. 223:163-166(1990).
CC -!- FUNCTION: Ice nucleation proteins enable bacteria to nucleate
crystallization in supercooled water.
CC -!- SUBCELLULAR LOCATION: Outer membrane (By similarity).
CC -!- DOMAIN: CONTAINS 153 IMPERFECT REPEATS OF THE CONSENSUS
REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.
CC -!- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE
NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.
CC -!- SIMILARITY: Belongs to the bacterial ice nucleation protein
family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X52970; CAA37140.1; -.
DR HSSP; P06620; LINA.
DR InterPro; IPR000258; Ice_nucleatn.
DR Pfam; PF00818; Ice_nucleation; 81.
DR PRINTS; PR00327; ICENUCLEATN.
DR PROSITE; PS00314; ICE_NUCLEATION; 57.
KW Ice nucleation; Repeat; Outer membrane.
SQ SEQUENCE 1567 AA; 152548 MW; C8B451D959ECAD63 CRC64;

Query Match 11.5%; Score 89.5; DB 1; Length 1567;
Best Local Similarity 26.4%; Pred. No. 7.2;
Matches 47; Conservative 23; Mismatches 59; Indels 49; Gaps 12;

QY 14 VGSALAG-----VFPQWG--GGNHN-----GGNNSGPDSTLSIYQGSNAALAL 59
Db 205 VYGSTLTGADOSRLVAGYGSTETAGDSDLIAGYGSTGTAGSDSSI-LAGYGSTQTAGR 263
QY 60 Q-----SDARKSETTITQSGY-----NGAD-----YDQLVTRVVTHEMAHFRNAT 102
Db 264 STLTAGYGSTQTAGBSRLTSGYSTATSGSDSAVISGYGSTQTAGSESSLTAGYGSTQT 323
QY 103 IDQWNAKNSDITVGYG-----GNNAALV-----NOTASDSSVMVROVGFNNATANQ 150
Db 324 A----RKGSIDITAG-YGSTGTAGSDSALIAGYGSTQTAGSESSLT--AGYGSTQTARK 374
```

```
RESULT 9
ICEN_PANAN
ID ICEN_PANAN STANDARD; PRT; 1034 AA.
AC Q47879;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ice nucleation protein inau.
GN INAU.
OS Pantoea ananas (Erwinia uredovora).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pantoea.
OX NCBI_TaxID=553;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KUIN-3;
RX MEDLINE=94264407; PubMed=7764866;
RA Michigami Y., Watabe S., Abe K., Obata H., Arai S.;
RT "Cloning and sequencing of an ice nucleation active gene of Erwinia
uredovora.";
RL Biosci. Biotechnol. Biochem. 58:762-764(1994).
CC -!- FUNCTION: Ice nucleation proteins enable bacteria to nucleate
crystallization in supercooled water.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- DOMAIN: CONTAINS IMPERFECT REPEATS OF A CONSENSUS OCTAPEPTIDE
PERIODICITY IS SUPERIMPOSED.
CC -!- G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A REGIONAL 48-RESIDUE
MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE
NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.
CC -!- SIMILARITY: Belongs to the bacterial ice nucleation protein
family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D14992; BAA03636.1; -.
DR PIR; JC2143; JC2143.
DR HSSP; P06620; LINA.
DR InterPro; IPR000258; Ice_nucleatn.
DR Pfam; PF00818; Ice_nucleation; 51.
DR PRINTS; PR00327; ICENUCLEATN.
DR PROSITE; PS00314; ICE_NUCLEATION; 34.
KW Ice nucleation; Repeat; Outer membrane.
FT DOMAIN 162..993 OCTAPEPTIDE PERIODICITY.
SQ SEQUENCE 1034 AA; 103378 MW; FA222523D333EADD CRC64;

Query Match 11.3%; Score 87.5; DB 1; Length 1034;
Best Local Similarity 25.2%; Pred. No. 6.6;
Matches 41; Conservative 18; Mismatches 41; Indels 63; Gaps 9;

QY 12 IVVSGSALAGVVPQW--GGGNHNGGNS-----GPDSTLSIYQGSNAAL 57
Db 161 IATYGSTLTGTHQSOLIAGYGSTETAGDSSTLIAGYGSTGTAGSDSTL-VAGYGSTQTA- 218
QY 58 ALQSDARKSETTITQSGYCGNGADYDQLVTRVVTHEMAHFRNATIDQWNAKNSDITVQ 117
Db 219 -----GEESQWAGYS-----TGTGKGSDLTAG- 243
QY 118 YG-----GNNAALV-----NOTASDSSVMVROVGFNNATANQ 150
Db 244 YGSTGTAGDSSLIAGYGSTQTAGDSSLT--AGYGSTQTACK 284

RESULT 10
ICEN_ERWHE
ID ICEN_ERWHE STANDARD; PRT; 1258 AA.
AC P16239;
```

```

DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-APR-2001 (Rel. 40, Last annotation update)
DE ICE nucleation protein.
GN ICEE.

OS Erwinia herbicola.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pantoea.
OX NCBI_TaxID=549;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=M1;
RX MEDLINE=90152370; PubMed=2515997;
RA Warren G.J., Corotto L.V.;
RT "The consensus sequence of ice nucleation proteins from Erwinia
RT herbicola, Pseudomonas fluorescens and Pseudomonas syringae.";
RL Gene 85:239-242(1989).
CC -!- FUNCTION: Ice nucleation proteins enable bacteria to nucleate
CC crystallization in supercooled water.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- DOMAIN: CONTAINS 126 IMPERFECT REPEATS OF A CONSENSUS OCTAPEPTIDE
CC A-G-Y-G-S-T-A-T; FURTHER ON A 16-RESIDUE AND A REGIONAL 48-RESIDUE
CC PERIODICITY IS SUPERIMPOSED.
CC -!- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE
CC NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.
CC -!- SIMILARITY: Belongs to the bacterial ice nucleation protein
CC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; W26382; AAA24823.1; -.
CC PIR; JQ0188; JQ0188.
CC DR HSP; P06620; IINA.
CC InterPro; IPR000258; Ice nucleatn.
CC Pfam; PF00818; Ice nucleation; 65.
CC PRINTS; PR00327; ICNUCLEATN.
CC PROSITE; PS00314; ICE_NUCLEATION; 45.
CC Ice nucleation; Repeat; Outer membrane.
CC DOMAIN 162 1217 OCTAPEPTIDE PERIODICITY.
CC SEQUENCE 1258 AA; 125084 MW; 59088A130077FBD4 CRC64;

Query Match 11.3%; Score 87.5; DB 1; Length 1258;
Best Local Similarity 25.2%; Pred. No. 8.2; Indels 53; Gaps 9;
Matches 41; Conservative 18; Mismatches 41;

QY 12 IVVSGSALAGVVPWV-GGGNNHGGNNS-----GPDSTLSIYQVGSANAAL 57
Db 161 IATYGSTLSGTHQSOLIAGYGSTAGDSSTLIAGYGSTGTAGADSTL-VAGYGSTQTA- 218
QY 58 ALQSDARKSETTITQSGYNGADYDQLVTRVVTTHMAHAFNNATIDQWNAKNSDITV 117
Db 219 -----GEESQVAGYGS-----TGTGKSGDLTAG- 243
QY 118 YG-----GNNAALV-----NOTASDSSVMVRQVGFNNATANQ 150
Db 244 YGSTGTAGDSSLIAGYGSTGTAGDSSLT--AGYGSTQTAQK 284

RESULT 11
CSGB SALT1
ID CSGB SALT1 STANDARD; PRT; 151 AA.
AC Q827M3;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Minor curlin subunit precursor.
GN CSGB OR STY1180 OR TI777.

OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahinia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagals K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrall B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=22931367; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodyanani V., Schwartz D.C., Blattner F.R.;
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RT and CT18.";
RL J. Bacteriol. 185:2330-2337(2003).
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
CC CURLIN MONOMERS.
CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL627269; CAD08267.1; -.
CC DR EMBL; AB018640; AAC069400.1; -.
CC KW Fimbrin; Signal; Complete proteome.
CC FT SIGNAL 1 21 POTENTIAL.
CC FT CHAIN 22 151 MINOR CURLIN SUBUNIT.
CC SQ SEQUENCE 151 AA; 16254 MW; 161C54326E573495 CRC64;

Query Match 11.2%; Score 86.5; DB 1; Length 151;
Best Local Similarity 26.2%; Pred. No. 0.89; Indels 37; Gaps 4;
Matches 34; Conservative 17; Mismatches 42;

QY 10 AAIVSGSALAGVVPWVGNNHGGNNSGPDSTLSIYQVGSANAALQSDARKSETT 69
Db 58 ARVQSGSKLLSVISQ--EGENNRKAVQDQAGNYFAYIEQTGNAN-----DAS 103
QY 70 ITQSGYNGADYDQLVTRVVTTHMAHAFNNATIDQWNAKNSDITVQYGGNNAALVQ 129
Db 104 ISQAYNGSA-----AIIQKSGNKANIT--QYGTQKTAVVQVK 140
QY 130 ASDSSVMVRQ 139
Db 141 QSHMAIRVTQ 150

RESULT 12
GP63 LEIDO
ID GP63 LEIDO STANDARD; PRT; 590 AA.
AC P232Z3;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)

```

RESULT 13	GP63 LEIME	STANDARD;	PRT;	646 AA.
ID	_GP63 LEIME			
AC	P43150;			
DT	01-NOV-1995 (Rel. 32, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Leishmanolysin C1 precursor (EC 3.4.24.36) (Cell surface protease)			
DE	(Major surface glycoprotein) (GP63 protein) (Promastigote surface			
DE	endopeptidase).			
GN	GP63-C1			
OS	Leishmania mexicana.			
OC	Eukaryota; Euzoenzoa; Kinetoplastida; Trypanosomatidae; Leishmania.			
OC	NCBI_TaxId=5665;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=MNYC/BZ/62/M379;			
RC	MEDLINE=93149206; PubMed=8426614;			
RA	Madina-Acosta E.; Kates R.E.; Russell D.G.;			
RA	"Structurally distinct genes for the surface protease of Leishmania			
RT	mexicana are developmentally regulated.";			
RL	Mol. Biochem. Parasitol. 57:31-46(1993).			
CC	-1- FUNCTION: Has an integral role during the infection of macrophages			
CC	in the mammalian host.			
CC	-1- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and			
CC	P1', and basic residues at P2 and P3'. A model nonapeptide is			
CC	cleaved at -Ala-Tyr- -Leu-Lys-Lys-.			
CC	-1- COFACTOR: Binds 1 zinc ion per subunit (By similarity).			
CC	-1- DEVELOPMENTAL STAGE: Expressed in both the promastigote and the			
CC	amastigote forms.			
CC	-1- SIMILARITY: Belongs to peptidase family M8.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation			
CC	at the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; X64394; CAA45733.1; -.			
DR	PIR; S19916; S19916.			
DR	HSP; P08148; ILML.			
DR	NEP; M08.001; -.			
DR	GlycoSuiteDB; P43150; -.			
DR	InterPro; IPR006025; Pept_M_Zn_Bs.			
DR	InterPro; IPR001577; Peptidase_M8.			
DR	Pfam; PF01457; Peptidase_M8; 1.			
DR	PRINTS; PR00782; LSHMANOLYSIN.			
DR	PROSITE; PS00442; ZINC_PROTEASE; 1.			
KW	Hydrolase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;			
KW	Zymogen; Signal; Cell adhesion; Multigene family.			
FT	SIGNAL 1 39			
FT	POTENTIAL			
FT	PROPEP 40 102			
FT	ACTIVATION PEPTIDE (POTENTIAL).			
FT	CHAIN 103 646			
FT	METAL 266 266			
FT	ZINC (CATALYTIC) (BY SIMILARITY).			
FT	ACT_SITE 267 267			
FT	BY SIMILARITY.			
FT	METAL 270 270			
FT	ZINC (CATALYTIC) (BY SIMILARITY).			
FT	METAL 336 336			
FT	ZINC (CATALYTIC) (BY SIMILARITY).			
FT	DISULFID 127 144			
FT	BY SIMILARITY.			
FT	DISULFID 193 232			
FT	BY SIMILARITY.			
FT	DISULFID 316 388			
FT	BY SIMILARITY.			
FT	DISULFID 395 458			
FT	BY SIMILARITY.			
FT	DISULFID 408 427			
FT	BY SIMILARITY.			
FT	DISULFID 417 492			
FT	BY SIMILARITY.			
FT	DISULFID 469 513			
FT	BY SIMILARITY.			
FT	DISULFID 518 568			
FT	BY SIMILARITY.			
FT	DISULFID 538 561			
FT	CARBOHYD 86 86			
FT	N-LINKED (GLCNAC. .) (POTENTIAL).			
FT	CARBOHYD 297 297			
FT	N-LINKED (GLCNAC. .) (POTENTIAL).			
FT	CARBOHYD 399 399			
FT	N-LINKED (GLCNAC. .) (POTENTIAL).			
FT	CARBOHYD 409 409			
FT	N-LINKED (GLCNAC. .) (POTENTIAL).			

FT	DISULFID	408	427	BY SIMILARITY.
FT	DISULFID	417	492	BY SIMILARITY.
FT	DISULFID	469	513	BY SIMILARITY.
FT	DISULFID	468	568	BY SIMILARITY.
FT	DISULFID	538	561	BY SIMILARITY.
FT	CARBOHYD	86	86	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	297	297	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	399	399	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	409	409	N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 433 433 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 445 445 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 456 456 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 501 501 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 646 AA; 69054 MW; FE48DDC78C1080A CRC64;

Query Match 10.9%; Score 84.5; DB 1; Length 646;
 Best Local Similarity 36.5%; Pred. No. 6.9;
 Matches 23; Conservative 7; Mismatches 12; Indels 21; Gaps 2;

QY 81 YDQLVTRVTHMAHAFRNAT-----IDQNAKNSDITV-----GOYG 119
 256 YDQLVTRVTHMAHAFRNATGFFGAGVIGVQVPHLRKDFNVSVITSTVVAAREQYG 315
 DB 120 GNN 122
 316 CNS 318

RESULT 14
 GP63_LEICH STANDARD; PRT; 599 AA.
 AC P15706;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Leishmanolysin precursor (BC 3.4.4.36) (Cell surface protease)
 DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface
 DE endopeptidase).
 GN GP63.
 OS Leishmania chagasi.
 OC Eukaryota; Eulenzozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 OX NCBI_TaxID=44271;
 [1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=90205976; PubMed=2320059;
 RA Ramamoorthy R., Donelson J.E., Paetz K.E., Maybodi M., Roberts S.C.,
 Wilson M.E.;
 RT "Three distinct RNAs for the surface protease gp63 are differentially
 RT expressed during development of Leishmania donovani chagasi
 RT Promastigotes to an infectious form."
 RL J. Biol. Chem. 267:1888-1895(1992).
 RN Mcl. Biochem. Parasitol. 39:267-274(1990).
 [2]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=92112918; PubMed=1370484;
 RA Ramamoorthy R., Donelson J.E., Paetz K.E., Maybodi M., Roberts S.C.,
 Wilson M.E.;
 RT "Three distinct RNAs for the surface protease gp63 are differentially
 RT expressed during development of Leishmania donovani chagasi
 RT Promastigotes to an infectious form."
 RL J. Biol. Chem. 267:1888-1895(1992).
 CC -!- FUNCTION: Has an integral role during the infection of macrophages
 CC in the mammalian host.
 CC -!- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and
 CC P1', and basic residues at P2 and P3'. A model nonapeptide is
 CC cleaved at -Ala-Tyr|-Leu-Lys-Lys-
 CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
 CC -!- SIMILARITY: Belongs to peptidase family M8.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL; M30672; AAA29238.1; -
 CC EMBL; M28527; AAA29235.1; -
 CC FIR; A44951; A44951.
 CC HSSP; P08148; 1LMML.
 CC MEROPS; M08.001; -
 CC InterPro; IPR006025; Pept_M_Zn_BS.
 CC InterPro; IPR001577; Peptidase_M8.

DR PFam; PF01457; Peptidase_M8; 1.
 DR PRINTS; PRO0782; LSHMANOLYSIN.
 DR PROSITE; PS00142; ZINC_PROTEASE; 1.
 KW Hydrolase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;
 KW Zymogen; Signal; Cell adhesion; GPI-anchor; Lipoprotein.
 FT SIGNAL 1 39 POTENTIAL.
 FT PROPEP 40 97 ACTIVATION PEPTIDE.
 FT CHAIN 98 574 LEISHMANOLYSIN.
 FT PROPEP 575 599 REMOVED IN NATURE FORM (BY SIMILARITY).
 FT METAL 261 261 ZINC (CATALYTIC) (BY SIMILARITY).
 FT ACT_SITE 262 262 BY SIMILARITY.
 FT METAL 265 265 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 331 331 ZINC (CATALYTIC) (BY SIMILARITY).
 FT DISULFID 122 139 BY SIMILARITY.
 FT DISULFID 188 227 BY SIMILARITY.
 FT DISULFID 311 383 BY SIMILARITY.
 FT DISULFID 390 452 BY SIMILARITY.
 FT DISULFID 403 422 BY SIMILARITY.
 FT DISULFID 412 486 BY SIMILARITY.
 FT DISULFID 463 507 BY SIMILARITY.
 FT DISULFID 512 562 BY SIMILARITY.
 FT DISULFID 532 585 BY SIMILARITY.
 FT CARBOHYD 297 297 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 394 394 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT LIPID 574 574 GPI-anchor amidated asparagine (By
 FT similarity).
 SQ SEQUENCE 599 AA; 63848 MW; 746730AE8E2A3E7C CRC64;

Query Match 10.8%; Score 84; DB 1; Length 599;
 Best Local Similarity 37.5%; Pred. No. 7;
 Matches 24; Conservative 5; Mismatches 29; Indels 6; Gaps 1;

QY 39 SGPDSTLSVQYGSANAALA-----LQSDARKSETTITQSGYNGADYDQLVTRVTHE 92
 203 SNTDFVMYVASVPSBEGVLAWATTCQVFSGDHPGVGINIPRAANTASRYDQLVTRVTHE 262
 QY 93 MAHA 96
 263 MAHA 266

RESULT 15
 HLXA_SERMA STANDARD; PRT; 1608 AA.
 AC P15320;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 01-NOV-1990 (Rel. 16, Last annotation update)
 DE Hemolysin precursor.
 GN SHLA.
 OS Serratia marcescens.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Serratia.
 OX NCBI_TaxID=615;
 [1]
 RN SEQUENCE FROM N.A., AND SEQUENCE OF 31-40.
 RP STRAIN=SN8;
 RC MEDLINE=88257037; PubMed=3290200;
 RA Poole K., Schiebel E., Braun V.;
 RT "Molecular characterization of the hemolysin determinant of Serratia
 RT marcescens."
 RL J. Bacteriol. 170:3177-3188(1988).
 CC -!- FUNCTION: Bacterial hemolysins are exotoxins that attack blood
 CC cell membranes and cause cell rupture by mechanisms not clearly
 CC defined.
 CC -!- FUNCTION: CELL-BOUND HEMOLYSIN, WHICH RELEASES HEME-IRON FROM
 CC ERYTHROCYTES BY INTERACTION WITH THE ERYTHROCYTE MEMBRANE. SHLA
 CC REQUIRES SHLB FUNCTION.
 CC -!- SUBCELLULAR LOCATION: Outer membrane.
 CC -!- SIMILARITY: TO P.MIRABILIS HEMOLYSIN (HPWA).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (C) 1993 - 2004 CompuGen Ltd.
OM protein - protein search, using sw model
Run on: August 2, 2004, 14:39:17 ; Search time 29.7 Seconds
(without alignments)
1604.150 Million cell updates/sec

Title: US-09-543-407-28
Perfect score: 775
Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVQVGFGNATANQY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mic:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	674	87.0	152	2	Q33802
2	566.5	73.1	150	2	Q7X243
3	530	68.4	149	2	Q7X240
4	494.5	62.5	152	16	Q8CW63
5	411.5	53.1	150	2	Q7X237
6	295	38.1	76	2	O54069
7	122	15.7	29	2	Q8S335
8	112	14.5	130	16	Q89J14
9	111.5	14.4	3501	16	Q8Y106
10	111.5	14.4	3552	16	Q8XSD6
11	110	14.2	139	16	Q8EIH3
12	108	13.9	1422	16	Q8EFU3
13	107	13.8	2174	16	Q92J08
14	106.5	13.7	502	16	Q8EIH4
15	99.5	12.8	171	16	Q89J13
16	99	12.8	714	16	Q7U5X6

17	98	12.6	179	2	Q33801
18	97.5	12.6	154	16	Q89J15
19	97	12.5	1748	5	Q84821
20	95	12.3	368	16	Q8EW56
21	95	12.3	400	10	Q8L4R8
22	95	12.3	1410	16	Q8CWJ0
23	94.5	12.2	160	16	Q8CW64
24	94	12.1	362	16	Q8EV84
25	94	12.1	645	16	Q7U1C5
26	94	12.1	646	16	O53818
27	93.5	12.1	644	5	Q8NNY9
28	93	12.0	157	16	Q88HG0
29	92.5	11.9	348	13	Q93397
30	92.5	11.9	644	5	O43994
31	92.5	11.9	1765	16	Q7V8S5
32	92	11.9	598	5	Q25275
33	91	11.7	152	2	Q7X241
34	91	11.7	573	10	Q3SAF2
35	91	11.7	614	16	Q7TYR8
36	91	11.7	615	16	P95249
37	91	11.7	1408	16	Q8E833
38	90.5	11.7	1286	2	Q841Y5
39	89.5	11.5	151	2	Q7X238
40	89.5	11.5	151	16	O7UC21
41	89.5	11.5	153	16	Q89J16
42	89.5	11.5	160	16	Q83RU7
43	89.5	11.5	1333	16	Q8PD38
44	89.5	11.5	2411	16	Q8PF72
45	89	11.5	354	16	P95248

ALIGNMENTS

RESULT 1

O33802

ID O33802 PRELIMINARY; PRT; 152 AA.

AC O33802; (TREMBLrel. 05, Created)

DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)

DT 01-JAN-1998 (TREMBLrel. 19, Last annotation update)

DE Agfa protein (Fragment).

GN AGFA.

OS Salmonella typhimurium.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Salmonella.

OX NCBI_TaxID=602;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=98053981; PubMed=9393832;

RA Sukupolvi S.S., Lorentz R.G., Gordon J.I., Bian Z., Pfeifer J.D.,

RA Normark S.J., Rhen M.;

RT "Expression of thin, aggregative fimbriae promotes interaction of

RT Salmonella typhimurium SR-11 with mouse small intestinal epithelial

RT cells.";

RL Infect. Immun. 65:5320-5325(1997).

DR EMBL; AJ000514; CAA04151.1;

FT NON_TER 152

SQ SEQUENCE 152 AA; 15401 MW; 9DA7DADC2364B006 CRC64;

Query Match 87.0%; Score 674; DB 2; Length 152;

Best Local Similarity 88.7%; Pred. No. 1.1e-47;

Matches 134; Conservative 2; Mismatches 15; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGSGGPDSTLSIYQGSANALALQ 60

Db 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGSGGPDSTLSIYQGSANALALQ 60

QY 61 SPARKSETTITQSGYNGGADYDQLVTRVVVTHEMAHAFRNATIDQWNAKNSDITVQYGG 120

Db 61 SPARKSETTITQSGYNGGADYDQLVTRVVVTHEMAHAFRNATIDQWNAKNSDITVQYGG 120

QY 121 NNAALVNTQASDSSVMVQVGFGNATANQY 151

```

Db 121 NNAALVNQTASDSSVMVVRQVGFNNATANQY 151
|||||
RESULT 2
Q7X243 ID Q7X243 PRELIMINARY; PRT; 150 AA.
AC Q7X243;
DT 01-OCT-2003 (TReMBLrel. 25, Created)
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DE Curlin-csgA protein.
GN CSGA.
OS Citrobacter sp. Fec2.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Citrobacter.
OX NCBI_TaxID=213763;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=Fec2;
RA Zogaj X., Bokranz W., Nimtz M., Romling U.;
RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RT Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158(2003).
DR EMBL; AJ515700; CAD56672.1; -
SQ SEQUENCE 150 AA; 15016 MW; 1D7141B8D6973DC6 CRC64;

Query Match 73.1%; Score 566.5; DB 2; Length 150;
Best Local Similarity 76.8%; Pred. No. 6.5e-39;
Matches 116; Conservative 10; Mismatches 24; Indels 1; Gaps 1;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHGGNSGPGDSTLSIYQYGSANAALQ 60
|||||
Db 1 MKLLOVAFAAIVVSGSALAGVVPQWGGGNGHGGNSGPGDSTLSIYQYGSANAALQ 59
|||||
QY 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFRNATIDQWNAKNSDITVGYGG 120
|||||
Db 60 SDARKSDTTIHQNGFGNGADVQGGSDNSTDLTQNGFGNNATIDQWNGKNSDITVGYGG 119
|||||
QY 121 NNAALVNQTASDSSVMVVRQVGFNNATANQY 151
|||||
Db 120 HNAALVNQTASDSSVLVHQVGFNNATANQY 150
|||||

RESULT 3
Q7X240 ID Q7X240 PRELIMINARY; PRT; 149 AA.
AC Q7X240;
DT 01-OCT-2003 (TReMBLrel. 25, Created)
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DE Curlin-csgA protein.
GN CSGA.
OS Citrobacter freundii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Citrobacter.
OX NCBI_TaxID=546;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=Fec4;
RA Zogaj X., Bokranz W., Nimtz M., Romling U.;
RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RT Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158(2003).
DR EMBL; AJ515701; CAD56675.1; -
SQ SEQUENCE 149 AA; 15260 MW; 946DD52017F648FD CRC64;

Query Match 68.4%; Score 530; DB 2; Length 149;
Best Local Similarity 71.5%; Pred. No. 6.3e-36;
Matches 108; Conservative 16; Mismatches 25; Indels 2; Gaps 1;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHGGNSGPGDSTLSIYQYGSANAALQ 60
|||||

```

```

Db 1 MKLLKVAFAAIVVSGSALAGVVPQW--CGNHGGSGSYGPPSSLSIYQYGSNNSANAQ 58
QY 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFRNATIDQWNAKNSDITVGYGG 120
|||||
Db 59 SDARKSDVTITQHGNGAVVGQGADDSITSLKQTGFQNSATIDQWNAKNAIDISVTQFG 118
|||||
QY 121 NNAALVNQTASDSSVMVVRQVGFNNATANQY 151
|||||
Db 119 RINGALVNQTASDSSNVLIIQVGFNNATANQH 149
|||||

RESULT 4
Q8CW63 ID Q8CW63 PRELIMINARY; PRT; 152 AA.
AC Q8CW63;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DE Major curlin subunit precursor.
GN CSGA OR C1306.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
DR EMBL; AB016759; AAN79779.1; -
KW Complete proteome.
SQ SEQUENCE 152 AA; 15064 MW; 3BA57F34C1240E83 CRC64;

Query Match 62.5%; Score 484.5; DB 16; Length 152;
Best Local Similarity 66.4%; Pred. No. 3.4e-32;
Matches 101; Conservative 15; Mismatches 35; Indels 1; Gaps 1;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQW--CGGNGHGGNSGPGDSTLSIYQYGSANAALQ 59
|||||
Db 1 MKLLKVAFAAIVVSGSALAGVVPQYGGGGNGHGGNSGPNSELNIYQYGGNSALAQ 60
|||||
QY 60 QSDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFRNATIDQWNAKNSDITVGYGG 119
|||||
Db 61 QADARNSLTITQHGNGADVQGGSDNSTDLTQNGFGNSATLDQWNGKSDITMTVQFG 120
|||||
QY 120 GNNALVNQTASDSSVMVVRQVGFNNATANQY 151
|||||
Db 121 GNGHADVQTASNSSVNTVQVGFNNATAHQY 152
|||||

RESULT 5
Q7X237 ID Q7X237 PRELIMINARY; PRT; 150 AA.
AC Q7X237;
DT 01-OCT-2003 (TReMBLrel. 25, Created)
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DE Curlin-csgA protein.
GN CSGA.
OS Enterobacter sakazakii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Enterobacter.
OX NCBI_TaxID=28141;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=Fec39;
RA Zogaj X., Bokranz W., Nimtz M., Romling U.;

```



```

RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RT Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158 (2003).
DR EMBL; AJ515702; CAD56678.1; -.
SQ SEQUENCE 150 AA; 15112 MW; 5DBB2D872DF15F3 CRC64;

Query Match 53.1%; Score 411.5; DB 2; Length 150;
Best Local Similarity 57.6%; Pred. No. 3.1e-26;
Matches 87; Conservative 25; Mismatches 38; Indels 1; Gaps 1;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSSGPDSTLSIYQYGSANAALALQ 60
DQ :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MKFIKVAALAAIVVSGSAGMAGMIQ-GGNGHGHGGYGGPNSFLNIYQNGGNSALALQ 59
QY 61 SDARKSETTTTQSYGNGADYDQLVTRVTHEMAHAFRNATIDQWNAKNSDITVGOYGG 120
DQ :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 60 TDAENSVLNISQTGGNGADYGGQSDSSINLTQNGFGNSATLDQWNSKDSVMNVSYQGG 119
QY 121 NNAALVNQTSDDSSVMVQVGFGNATANOY 151
DQ :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 120 LNALVDQTSNSTVNTQIGFGNHATAHQY 150

RESULT 6
Q54069 PRELIMINARY; PRT; 76 AA.
AC Q54069
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE SEF17 fimbria (fragment).
GN AGFA.
OS Salmonella enteritidis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=592;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SE30;
RA Cox J.M., Eglezos S., Woolcock J.B.;
RT "Virulence of Salmonella enteritidis in chickens correlates with
RT colony morphology and expression of SEF17 fimbriae.";
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U53207; AAA98671.1; -.
FT NON_TER 1 76
FT NON_TER 76 76
SQ SEQUENCE 76 AA; 7704 MW; 2FD5411241A7BCB1 CRC64;

Query Match 38.1%; Score 295; DB 2; Length 76;
Best Local Similarity 78.9%; Pred. No. 4.7e-17;
Matches 60; Conservative 1; Mismatches 15; Indels 0; Gaps 0;

QY 30 GNHGGGSSGPDSTLSIYQYGSANAALALQSDARKSETTTTQSYGNGADYDQLVTRVV 89
DQ :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1 GNHGGGSSGPDSTLSIYQYGSANAALALQSDARKSETTTTQSYGNGADYDQGDNST 60
QY 90 THEMAHAFRNATIDQ 105
DQ :|||||:
Db 61 IELTQNGFRNATIDQ 76

RESULT 7
Q54069 PRELIMINARY; PRT; 29 AA.
AC Q54069
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE Curlin subunit monomer (fragment).
GN CSGA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.

```

```

OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC TRANSPOSON-Insertion sequence ISI;
RX MEDLINE=99314153; PubMed=10386375;
RA La Ragione R.M., Collighan R.J., Woodward M.J.;
RT "Non-curlation of Escherichia coli O78:K80 isolates associated with
RT ISI insertion in csfB and reduced persistence in poultry infection.";
RL FEMS Microbiol. Lett. 175:247-253 (1999).
DR EMBL; AJ131756; CAB45380.1; -.
FT NON_TER 29
FT NON_TER 29 29
SQ SEQUENCE 29 AA; 2789 MW; E290DFC07ABBB243 CRC64;

Query Match 15.7%; Score 122; DB 2; Length 29;
Best Local Similarity 89.7%; Pred. No. 0.0021;
Matches 26; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGG 29
DQ :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1 MKLLKVAFAAIVVSGSALAGVVPQYGGG 29

RESULT 8
Q89JI4 PRELIMINARY; PRT; 130 AA.
AC Q89JI4
DT 01-JUN-2003 (TReMBLrel. 24, Created)
DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE BLI5299 protein.
GN BLI5299.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RX MEDLINE=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA Sasanoto S., Watanabe A., Idesawa K., Iriguchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimo S., Tsuruoka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110.";
RL DNA Res. 9:189-197 (2002).
DR EMBL; AP005954; BAC50564.1; -.
KW Complete proteome.
SQ SEQUENCE 130 AA; 12699 MW; ACFB2D66A48D260F CRC64;

Query Match 14.5%; Score 112; DB 16; Length 130;
Best Local Similarity 26.5%; Pred. No. 0.085;
Matches 40; Conservative 28; Mismatches 55; Indels 28; Gaps 6;

QY 4 LKVAFAAIVVSGSALAGVVPQWGGGNGHNGSSGPDSTLSIYQYGSANAALALQSDA 63
DQ :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1 MRITYLVATAIALSALTIVDAQ-----AGNSA-----SVLQFGTNSFSQTGS 45
QY 64 RKSETTTTQSYGNGADYDQLVTRVTHEMAHAFRNATIDQW----NAKNSDITVGOYG 119
DQ :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 46 TSNNATTLQFATNTA-----TTLQTSLLIV--NTAVTGGGTTATASNTALT-GQVG 96

Query Match 14.5%; Score 112; DB 16; Length 130;
Best Local Similarity 26.5%; Pred. No. 0.085;
Matches 40; Conservative 28; Mismatches 55; Indels 28; Gaps 6;

QY 120 GNNAALVNQTSDDSSVMVQVGFGNATANO 150
DQ :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 97 GNSSLIGQIGANTAGVQGLGILNGSTILQ 127

RESULT 9
Q89I06 PRELIMINARY; PRT; 3501 AA.
ID Q89I06
AC Q89I06
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)

```

```

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
GN RSC0887 OR RS06116.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Ariat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Chandler M., Choisme N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Sigulier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
DR EMBL; AL646061; CAD14589.1; -.
DR GO; GO:0004519; F:endonuclease activity; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001604; Endonuclease.
DR InterPro; IPR008638; Haemagg_act.
DR Pfam; PF05594; Fil_haemagg; 20.
DR Pfam; PF05960; Haemagg_act; 1.
DR PROSITE; PS01070; NUCLEASE_NON_SPEC; 1.
DR Complete proteome.
KW SEQUENCE 3501 AA; 348421 MW; 290B41C99018A107 CRC64;
SQ SEQUENCE 3501 AA; 348421 MW; 290B41C99018A107 CRC64;

Query Match 14.4%; Score 111.5; DB 16; Length 3501;
Best Local Similarity 29.3%; Pred. No. 5.2;
Matches 41; Conservative 20; Mismatches 52; Indels 27; Gaps 5;

QY 15 SCSALAGVVPQWGGGNGGG-NSSGPDSTLSIYQGSANAA----- 56
DB 2431 SCSHFTAGPAGMDLGRNVGGPNSSG---VGLAPYGSASADNAAGNSRQNASVVG 2486
QY 57 LALQSDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFNNATIDQWNAKNSDITVG 116
DB 2487 KSVQVQARTGDTIVSGSISALSDVLLAKQKVDIVA---GNDTSSR-HEHSDRTIG 2541
QY 117 QYGGNNAALVNQTASDSSVM 136
DB 2542 DLGGNGYSGTVGVSASSTL 2561

RESULT 10
QXASD6 PRELIMINARY; PRT; 3552 AA.
AC Q8XSD6;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DE Probable hemagglutinin-related protein.
GN RSP0540 OR RS06117.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OG Plasmid megaplasmid.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Ariat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Chandler M., Choisme N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Sigulier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
DR EMBL; AL646061; CAD14589.1; -.
DR GO; GO:0004519; F:endonuclease activity; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001604; Endonuclease.
DR InterPro; IPR008638; Haemagg_act.
DR Pfam; PF05594; Fil_haemagg; 20.
DR Pfam; PF05960; Haemagg_act; 1.
DR PROSITE; PS01070; NUCLEASE_NON_SPEC; 1.
DR Complete proteome.
KW SEQUENCE 3501 AA; 348421 MW; 290B41C99018A107 CRC64;
SQ SEQUENCE 3501 AA; 348421 MW; 290B41C99018A107 CRC64;

Query Match 14.4%; Score 111.5; DB 16; Length 3501;
Best Local Similarity 29.3%; Pred. No. 5.2;
Matches 41; Conservative 20; Mismatches 52; Indels 27; Gaps 5;

QY 15 SCSALAGVVPQWGGGNGGG-NSSGPDSTLSIYQGSANAA----- 56
DB 2431 SCSHFTAGPAGMDLGRNVGGPNSSG---VGLAPYGSASADNAAGNSRQNASVVG 2486
QY 57 LALQSDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFNNATIDQWNAKNSDITVG 116
DB 2487 KSVQVQARTGDTIVSGSISALSDVLLAKQKVDIVA---GNDTSSR-HEHSDRTIG 2541
QY 117 QYGGNNAALVNQTASDSSVM 136
DB 2542 DLGGNGYSGTVGVSASSTL 2561

RESULT 10
QXASD6 PRELIMINARY; PRT; 3552 AA.
AC Q8XSD6;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DE Probable hemagglutinin-related protein.
GN RSP0540 OR RS06117.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OG Plasmid megaplasmid.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Ariat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Chandler M., Choisme N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Sigulier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
DR EMBL; AL646061; CAD14589.1; -.
DR GO; GO:0004519; F:endonuclease activity; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001604; Endonuclease.
DR InterPro; IPR008638; Haemagg_act.
DR Pfam; PF05594; Fil_haemagg; 20.
DR Pfam; PF05960; Haemagg_act; 1.
DR PROSITE; PS01070; NUCLEASE_NON_SPEC; 1.
DR Complete proteome.
KW SEQUENCE 3501 AA; 348421 MW; 290B41C99018A107 CRC64;
SQ SEQUENCE 3501 AA; 348421 MW; 290B41C99018A107 CRC64;

Query Match 14.4%; Score 111.5; DB 16; Length 3552;
Best Local Similarity 29.3%; Pred. No. 5.3;
Matches 41; Conservative 20; Mismatches 52; Indels 27; Gaps 5;

QY 15 SCSALAGVVPQWGGGNGGG-NSSGPDSTLSIYQGSANAA----- 56
DB 2430 SCSHFTAGPAGMDLGRNVGGPNSSG---VGLAPYGSASADNAAGNSRQNASVVG 2485
QY 57 LALQSDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFNNATIDQWNAKNSDITVG 116
DB 2486 KSVQVQARTGDTIVSGSISALSDVLLAKQKVDIVA---GNDTSSR-HEHSDRTIG 2540
QY 117 QYGGNNAALVNQTASDSSVM 136
DB 2541 DLGGNGYSGTVGVSASSTL 2560

RESULT 11
Q8EIH3 PRELIMINARY; PRT; 139 AA.
AC Q8EIH3;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Minor curlin subunit CsgB, putative.
GN S00866.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.
OX NCBI_TaxID=70863;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NR-1;
RX MEDLINE=22297686; PubMed=12368813;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Meche B., Clayton R.A.,
RA Meyer R.T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
RA Desoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
RA Vamathevan J., Weidman J., Imprim M., Lee K., Berry K., Lee C.,
RA Mueller J., Khouri H., Gill J., Utterback T.R., McDonald L.A.,
RA Feldblyum T.V., Smith H.O., Venter J.C., Nealsen K.H., Fraser C.M.;
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis.";
RL Nat. Biotechnol. 20:1118-1123(2002).
DR EMBL; AE015532; AAN53942.1; -.
DR TIGR; S00866; -.
KW Complete proteome.
SQ SEQUENCE 139 AA; 14811 MW; 41EC1CFA76957920 CRC64;

Query Match 14.2%; Score 110; DB 16; Length 139;
Best Local Similarity 28.3%; Pred. No. 0.13;
Matches 30; Conservative 19; Mismatches 33; Indels 24; Gaps 3;

QY 39 SGPDPSTLSIYQGSANALQSDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFN 98
DB 41 SGRDNLIDLVOQGTANQIVFGSGSDNS-AYVQAGNDN---ISLVQIGT----- 87
QY 99 NNATIDQWNAKNSDITVGOYGGNNAALVNQTASDSSVMVQVGFN 144

```

RX	MEDLINE=21396508; PubMed=11491431;
RA	Finan T.M., Weidner S., Wong K., Buhrmaster J., Chain P.,
RA	Vorhoehter F.J., Hernandez-Lucas I., Becker A., Gouzy J.,
RA	Golding B., Puehler A.;
RT	"The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-
RT	fixing endosymbiont Sinorhizobium meliloti.";
RL	Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).
DR	EMBL; AL603645; CAC49389.1; -.
DR	PIN; E95965; E95965.
DR	GO; GO:0046821; C:extrachromosomal DNA; IEA.
DR	GO; GO:0004601; F:peroxidase activity; IEA.
DR	GO; GO:0006979; P:response to oxidative stress; IEA.
DR	InterPro; IPR005546; Autotransporter.
DR	InterPro; IPR002016; Peroxidase.
DR	InterPro; IPR002173; PfkB.
DR	Pfam; PF03797; Autotransporter; 1.
DR	PROSITE; PS00435; PEROXIDASE_1; 1.
DR	PROSITE; PS00583; PFKB_KINASES_1; 2.
KW	Plasmid; Hypothetical protein; Complete proteome.
SQ	SEQUENCE 2174 AA; 203314 MW; 008EB68297B44182 CRC64;

Query Match	13.8%; Score 107; DB 16; Length 2174;
Best Local Similarity	26.2%; Pred.No.6.8;
Matches	37; Conservative 21; Mismatches 61; Indels 22; Gaps 6

QY	11 AIIVSGSALAGVWPQ--WGCGGNHNGGNSGGPDTLSIYOYGSAHALAQSDAR---K 65
DB	693 AIAATAGAGVILAQSIGGGG--GGAATGGDAGFGSFQIGGGGGGYANTANVGFK 749
QY	66 SETTITGSYNGADYDQLVTRVTTHMAHAFRNNAITDOWNAK---NSDITVGOYGNN 122
DB	750 GLTLTTGGSHAAG-----IVAQSVGGGGTGCTASSYAGIGFTASVAVGTGGNG 800
QY	123 AA--LVNQTASDSVMVRQVG 141
DB	801 GAGGEVSVLTDIAIRTCQG 821

RESULT 14	
QSEIH4	
ID Q8EIH4	PRELIMINARY; PRT; 502 AA.
AC Q8EIH4	
DT 01-MAR-2003 (TrEMBLrel. 23, Created)	
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)	
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)	
DE Conserved hypothetical protein.	
DN SO0865	
OS Shewanella oneidensis.	
OC Bacteriä; Proteobacteria; Gammaproteobacteria; Alteromonadales;	
OC Alteromonadaceae; Shewanella.	
OX NCBI_taxid=70863;	
RN [1]	
RP SEQUENCE FROM N.A.	
RC STRAIN=MR-1;	
RX MEDLINE=23297686; PubMed=12368813;	
RA Heidelberg J.F., Paulsen I.T., Neilson K.E., Gaidos E.J., Nelson W.C.,	
RA Read T.D., Eisen J.A., Seshadi R., Ward N., Methe B., Clayton R.A.,	
RA Meyer R.T., Tsapin A., Scott J., Beaman M., Brinkac L., Daugherty S.,	
RA Madupu R., Dodson R.J., Durkin A.S., Hart D.H., Kolonay J.F.,	
RA Vamathevan J., Peterson J.D., Umayam L.A., White O., Wolf A.M.,	
RA Mueller J.V., Khouri H., Gill J., Utterback T.R., McDonald L.A.,	
RA Feilblyum T.V., Smith H.O., Venter J.C., Nealson K.H., Fraser C.M.;	
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium	
RT Shewanella oneidensis.";	
RL Nat. Biotechnol. 20:1118-1123(2002).	
DR EMBL; AE015532; AAN53941.1; -.	
DR TIGR; SO0865; -.	
KW Hypothetical protein; Complete proteome.	
SQ SEQUENCE 502 AA; 52441 MW; D08CA23D6C46B62D CRC64;	

Query Match	13.7%; Score 106.5; DB 16; Length 502;
Best Local Similarity	24.2%; Pred.No.1.2;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:35:42 ; Search time 44.9 Seconds
(without alignments)
950.215 Million cell updates/sec

Title: US-09-543-407-30
Perfect score: 768
Sequence: 1 MKLLKVAFAIAIVVSSALA.....DSSVMYRVQVFGNNATANQY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04:*

- 1: Geneseqp1980s:*
- 2: Geneseqp1990s:*
- 3: Geneseqp2000s:*
- 4: Geneseqp2001s:*
- 5: Geneseqp2002s:*
- 6: Geneseqp2003as:*
- 7: Geneseqp2003bs:*
- 8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	768	100.0	151	3	AAB36355	Aab36355 Agfa::PT3
2	700	91.1	151	3	AAB36352	Aab36352 Agfa::PT3
3	681	88.7	151	2	AAR74625	Aar74625 Agfa sequ
4	681	88.7	151	3	AAB36341	Aab36341 Salmonell
5	676	88.0	151	2	AAW23570	Aaw23570 Salmonell
6	659	85.8	151	3	AAB36347	Aab36347 Agfa::PT3
7	657	85.5	151	3	AAB36351	Aab36351 Agfa::PT3
8	613	79.8	151	3	AAB36354	Aab36354 Agfa::PT3
9	603	78.5	151	3	AAB36346	Aab36346 Agfa::PT3
10	600	78.1	151	3	AAB36353	Aab36353 Agfa::PT3
11	599	78.0	151	3	AAB36349	Aab36349 Agfa::PT3
12	597	77.7	151	3	AAB36350	Aab36350 Agfa::PT3
13	566	73.7	151	3	AAB36348	Aab36348 Agfa::PT3
14	517	67.3	151	3	AAB36343	Aab36343 Escherich
15	512	66.7	151	7	ABR84651	AbR84651 E. coli C
16	496	64.6	120	2	AAR62761	Aar62761 Agfa sequ
17	496	64.6	120	2	AAW23569	Aaw23569 Salmonell
18	439	57.2	142	2	AAW52664	Aaw52664 Fibronect
19	367	47.8	122	2	AAR52663	Aar52663 FNB curli
20	215	28.0	45	3	AAB36316	Aab36316 Salmonell
21	132	17.2	22	3	AAB36318	Aab36318 Salmonell
22	115	15.0	22	3	AAB36325	Aab36325 Salmonell
23	115	15.0	22	3	AAB36339	Aab36339 Salmonell
24	115	15.0	22	3	AAB36320	Aab36320 Salmonell
25	113	14.7	24	7	ABR83644	AbR83644 E. coli c

26	109	14.2	23	3	AAB36340	Aab36340 Salmonell
27	109	14.2	23	3	AAB36324	Aab36324 Salmonell
28	109	14.2	23	3	AAB36319	Aab36319 Salmonell
29	104.5	13.6	151	3	AAB36344	Aab36344 Escherich
30	102	13.3	26	7	ABR82649	AbR82649 E. coli V
31	101.5	13.2	520	6	AAO16497	Aao16497 Argiope t
32	99.5	13.0	151	3	AAB36342	Aab36342 Salmonell
33	98	12.8	26	7	ABR82645	AbR82645 E. coli c
34	96	12.5	19	3	AAB36323	Aab36323 Salmonell
35	96	12.5	19	3	AAB36336	Aab36336 Salmonell
36	96	12.5	19	3	AAB36328	Aab36328 Salmonell
37	95.5	12.4	908	4	ABB65007	Abb65007 Drosophil
38	95	12.4	718	3	AAV78364	Aav78364 H. pylori
39	94.5	12.3	908	4	ABB62355	Abb62355 Drosophil
40	94	12.2	354	7	ABO23520	AbO23520 Mycobacte
41	93.5	12.2	2204	6	ABU32640	Abu32640 Protein e
42	92.5	12.0	738	2	AAW56163	Aaw56163 New DNA S
43	92.5	12.0	850	4	ABB65764	Abb65764 Drosophil
44	92.5	12.0	1028	4	ABB62708	Abb62708 Drosophil
45	92.5	12.0	1721	6	ABU34624	Abu34624 Protein e

ALIGNMENTS

RESULT 1
AAB36355
ID AAB36355 standard; protein; 151 AA.
XX
AC AAB36355;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa::PT3#10 amino acid sequence SEQ ID NO:30.
XX
XX Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
XX vaccine; immune response; immunogen.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR N-PSDB; AAC64631.
XX
PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 139; 139pp; English.
XX
CC The present invention describes a recombinant agfa gene (1) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CSga and Agfa-homologue fimbriin subunits, respectively. (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native

CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;

Query Match 100.0%; Score 768; DB 3; Length 151;
 Best Local Similarity 100.0%; Pred. No. 1.7e-66;
 Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
 QY 61 SDARKSETTITQSGYNGADYVQGGADNSTIELTQGFNNATYDQLVTRVVTHEMAHAGG 120
 DB 61 SDARKSETTITQSGYNGADYVQGGADNSTIELTQGFNNATYDQLVTRVVTHEMAHAGG 120
 QY 121 NNAALVNQTASDSSVWVRQVGFNNATANQY 151
 DB 121 NNAALVNQTASDSSVWVRQVGFNNATANQY 151

RESULT 2
 AAB36352
 ID AAB36352 standard; protein; 151 AA.
 AC AAB36352;
 XX
 XX 26-FEB-2001 (first entry)
 XX
 XX Agfa::PT3#7 amino acid sequence SEQ ID NO:24.

DE Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 XX

OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.

XX WO200060102-A2.

XX 12-OCT-2000.

XX 05-APR-2000; 2000WO-CA000356.

XX 05-APR-1999; 99US-0127888P.

XX (UYVI-) UNIV VICTORIA.

XX White AP, Doran JL, Collison SK, Kay WW;

XX WPI; 2000-672631/65.

XX N-PSDB; AAC64628.

PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant agfa
 PT protein useful for eliciting immune response in animal.

XX Disclosure; Page 136; 139pp; English.
 PS

XX The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CSga and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;

Query Match 91.1%; Score 700; DB 3; Length 151;
 Best Local Similarity 89.9%; Pred. No. 6.8e-60;
 Matches 143; Conservative 0; Mismatches 0; Indels 16; Gaps 2;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
 QY 61 SDARKSETTITQSGYNGADYVQGGADNSTIELTQGFNNATYDQLVTRVVT 112
 DB 61 SDARKSETTITQSGYNGADYVQGGADNSTIELTQGFNNATIDQWNAKNYDQLVTRVVT 120
 QY 113 HEMAHAGGNNALVNQTASDSSVWVRQVGFNNATANQY 151
 DB 121 HEMAHA-----NQ TASDSSVWVRQVGFNNATANQY 151

RESULT 3
 AAR74625

ID AAR74625 standard; protein; 151 AA.

XX AAR74625;

XX 25-MAR-2003 (revised)

DT 26-JUN-1995 (first entry)

XX Agfa sequence.

XX Salmonella; Agfa; vaccine.

XX Salmonella.

XX WO9425598-A2.

XX 10-NOV-1994.

XX 26-APR-1994; 94WO-IB000207.

XX 26-APR-1993; 93US-00054452.

XX (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
 PA (KING/) KING J.

```

PI Kay WW, Collinson SK, Clouthier SC, Doran JL;
XX WPI; 1994-358275/44.
DR N-PSDB; AAO87467.
XX
XX Eliciting an immune response to Salmonella - using attenuated Salmonella
PT strains, vector constructs, or compens. contg. fimbrial type proteins.
XX
XX Disclosure; Fig 7B; 95pp; English.
XX
XX The Salmonella AgfA protein and DNA are used in vaccine and genetic
CC immunization compositions, respectively, to elicit an immune response to
CC Salmonella in animals (e.g. food producing animals) and humans. (Updated
CC on 25-MAR-2003 to correct PN field.)
XX
XX Sequence 151 AA;
XX
Query Match      88.7%; Score 681; DB 2; Length 151;
Best Local Similarity 90.7%; Pred. No. 4.7e-58;
Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;
QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATYDQLVTRVVTHEMAHAGG 120
DB 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
RESULT 4
AAB36341
ID AAB36341 standard; protein; 151 AA.
XX
AC AAB36341;
XX
DT 26-FEB-2001 (first entry)
XX
DE Salmonella enteritidis AgfA amino acid sequence SEQ ID NO:5.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbria; epitope;
KW vaccine; immune response; immunogen.
XX
OS Salmonella enteritidis.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
XX (UUVI-) UNIV VICTORIA.
XX
XX White AP, Doran JL, Collinson SK, Kay WW;
XX
XX WPI; 2000-672631/65.
XX
XX N-PSDB; AAC64617.
XX
XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 135; 139pp; English.
XX
XX The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:

```

```

CC (1) use of thin aggregative fimbriae (SEF17/TAP) nucleation depended
CC assembly system of strains of Salmonella. Escherichia coli and
CC enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbria subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
XX Sequence 151 AA;
XX
Query Match      88.7%; Score 681; DB 3; Length 151;
Best Local Similarity 90.7%; Pred. No. 4.7e-58;
Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;
QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATYDQLVTRVVTHEMAHAGG 120
DB 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
RESULT 5
AAW23570
ID AAW23570 standard; protein; 151 AA.
XX
AC AAW23570;
XX
DT 25-MAR-2003 (revised)
DT 29-SEP-1997 (first entry)
XX
DE Salmonella enteritidis 27655-3b agfA.
XX
KW Enteropathogenic bacteria; enterobacteria; S. enteritidis; antibody.
XX
OS Salmonella enteritidis.
XX
XX Key Location/Qualifiers
FT Misc-difference 123
FT /note= "Encoded by GCC"
XX
XX US5635617-A.
XX
XX 03-JUN-1997.
XX
XX 26-APR-1994; 94US-00233788.
XX
XX 26-APR-1993; 93US-00054452.
XX
XX (UUVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX
XX Collinson SK, Kay WW, Doran JL;

```

```

XX WPI; 1997-309886/28.
XX DR N-PSDB; AAT74142.
XX PT Isolated Salmonella gene agfa - used for diagnosis of Salmonella or
XX PT enteropathogenic bacteria of the Enterobacteria family.
XX PS Example 2; Fig 7; 85pp; English.
XX CC The present sequence represents agfa encoded by the full agfa gene
XX CC derived from Salmonella enteritidis 27655-3b. The nucleic acid can be
XX CC used to provide diagnostic assays for Salmonella and/or enteropathogenic
XX CC bacteria of the family Enterobacteria. It can also be used to provide
XX CC proteins and antibodies which can be used for assays. The nucleic acid
XX CC sequence can be used to provide probes or primers which can specifically
XX CC hybridize to nucleic acid molecules from greater than 99% of Salmonella
XX CC strains that are pathogenic to warm-blooded animals relative to nucleic
XX CC acid molecules from virtually all other microbial organisms. (Updated on
XX CC 25-MAR-2003 to correct PF field.)
XX SQ Sequence 151 AA;
    Query Match      88.0%; Score 676; DB 2; Length 151;
    Best Local Similarity 90.1%; Pred. No. 1.4e-57;
    Matches 136; Conservative 2; Mismatches 13; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSGPDSLTSLIYQYGSANAALALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSGPDSLTSLIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGQADNSTIETQTGFERNATYDQLVTRVVTHEMAHAGG 120
DB 61 SDARKSETTITQSGYNGADVGQADNSTIETQTGFERNATIDQWNAKNSDITVGQYGG 120
QY 121 NNAALVNOTASDSSVWVRQVGFGNNTANQY 151
DB 121 NNPALVNOTASDSSVWVRQVGFGNNTANQY 151
RESULT 6
AAB36347
ID AAB36347 standard; protein; 151 AA.
XX AC AAB36347;
XX DT 26-FEB-2001 (first entry)
XX DE Agfa::PT3#2 amino acid sequence SEQ ID NO:14.
XX KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
XX KW vaccine; immune response; immunogen.
XX OS Salmonella enteritidis.
XX OS Escherichia coli.
XX OS Synthetic.
XX PN WO200060102-A2.
XX PD 12-OCT-2000.
XX PF 05-APR-2000; 2000WO-CA000356.
XX PR 05-APR-1999; 99US-0127888P.
XX PA (UYVI-) UNIV VICTORIA.
XX PI White AP, Doran JL, Collison SK, Kay WW;
XX DR WPI; 2000-672631/65.
XX DR N-PSDB; AAC64623.
XX PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
XX PT which encodes foreign epitope or antigen, expresses recombinant Agfa

```

```

PT protein useful for eliciting immune response in animal.
XX Disclosure; Page 136; 139pp; English.
XX CC The present invention describes a recombinant agfa gene (I) where a
XX CC segment of the gene has been replaced by a segment of a foreign DNA
XX CC sequence which encodes a foreign epitope or antigen. Also described are:
XX CC (1) use of thin aggregative fimbriae (SEF17/IAF) nucleation depended
XX CC assembly system of strains of Salmonella, Escherichia coli and
XX CC Enterobacteriaceae for the production of fimbriae comprising recombinant
XX CC Agfa, CsgA and Agfa-homologue fimbrin subunits, respectively; (2)
XX CC directing recombination of a recombinant gene into the chromosome of the
XX CC homologous species; (3) directing recombination of a recombinant gene
XX CC back into the chromosome of the homologous species, replacing the native
XX CC copy of that gene; and (4) eliciting an immune response in an animal, Agfa
XX CC comprising separating an amino acid polymer comprising a recombinant Agfa
XX CC protein containing a replacement segment or segments of foreign amino
XX CC acid sequence or sequences grown on a Salmonella, E. coli or
XX CC Enterobacteriaceae host cell, from the host cell and introducing the
XX CC polymer into the animal in conjunction with a carrier or diluent. (I) is
XX CC useful for the expression of recombinant Agfa protein which is useful for
XX CC eliciting an immune response in an animal. In a fimbrial presentation
XX CC system the heterologous antigens are presented in high numbers (up to
XX CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
XX CC immunogenicity and adhesion properties relevant for an efficient live
XX CC vaccine, the carrier fimbrial subunit proteins are usually strong
XX CC immunogens, which may be important for directing an immune response
XX CC against the inserted epitope, and hybrid fimbriae are easy and
XX CC inexpensive to purify in large amount. The present sequence is given in
XX CC the exemplification of the present invention
XX SQ Sequence 151 AA;
    Query Match      85.8%; Score 659; DB 3; Length 151;
    Best Local Similarity 81.9%; Pred. No. 6.4e-56;
    Matches 136; Conservative 0; Mismatches 0; Indels 30; Gaps 2;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSGPDSLTSLIYQYGSANAALALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSGPDSLTSLIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGQADNSTIETQTGFERNAT-----YDQ 105
DB 61 SDARKSETTITQSGYNGADVGQADNSTIETQTGFERNATIDQWNAKNSDITVGQYDQ 120
QY 106 LVTRVVTHEMAHAGGNNALVNQTASDSVWVRQVGFGNNTANQY 151
DB 121 LVTRVVTHEMAHAGGNNALVNQTASDSVWVRQVGFGNNTANQY 151
RESULT 7
AAB36351
ID AAB36351 standard; protein; 151 AA.
XX AC AAB36351;
XX DT 26-FEB-2001 (first entry)
XX DE Agfa::PT3#6 amino acid sequence SEQ ID NO:22.
XX KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
XX KW vaccine; immune response; immunogen.
XX OS Salmonella enteritidis.
XX OS Escherichia coli.
XX OS Synthetic.
XX PN WO200060102-A2.
XX PD 12-OCT-2000.
XX PF 05-APR-2000; 2000WO-CA000356.
XX PT

```



```
PR 05-APR-1999; 99US-0127888P.
XX (UYVI-) UNIV VICTORIA.
XX White AP, Doran JL, Collison SK, Kay WW;
XX WPI; 2000-672631/65.
XX N-PSDB; AAC64627.
XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX Disclosure; Page 137; 139pp; English.
XX The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/7AF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating a replacement segment or segments of foreign amino
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX SQ Sequence 151 AA;
Query Match 85.5%; Score 657; DB 3; Length 151;
Best Local Similarity 81.9%; Pred. No. 1e-55;
Matches 136; Conservative 0; Mismatches 0; Indels 30; Gaps 2;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVCGQADNSTIETQNGFRNNATYDQLVTRVVTHEMAHA-- 118
DB 61 SDARKSETTITQSGYNGADVCGQADN-----YDQLVTRVVTHEMAHADQ 105
QY 119 -----GGNNAALVNQTASDSSVMVRQVGFNNATANQY 151
DB 106 WNKNKSDITVQYGGNNAALVNQTASDSSVMVRQVGFNNATANQY 151
RESULT 8
AAB36354
ID AAB36354 standard; protein; 151 AA.
XX AAB36354;
XX AAB36354;
XX 26-FEB-2001 (first entry)
XX AgfA::PT3#9 amino acid sequence SEQ ID NO:28.
XX Salmonella; agfA; chromosomal gene replacement; fimbrial; epitope;
XX vaccine; immune response; immunogen.
```

```
XX Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX WO200060102-A2.
XX 12-OCT-2000.
XX 05-APR-2000; 2000WO-CA000356.
XX 05-APR-1999; 99US-0127888P.
XX (UYVI-) UNIV VICTORIA.
XX White AP, Doran JL, Collison SK, Kay WW;
XX WPI; 2000-672631/65.
XX N-PSDB; AAC64630.
XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX Disclosure; Page 138; 139pp; English.
XX The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/7AF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX SQ Sequence 151 AA;
Query Match 79.8%; Score 613; DB 3; Length 151;
Best Local Similarity 74.6%; Pred. No. 1.9e-51;
Matches 123; Conservative 0; Mismatches 0; Indels 44; Gaps 2;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVCGQADNSTIETQNGFRNNATYDQLVTRVVTHEMAHA-- 118
DB 61 SDARKSETTITQSGYNGAD-----YDQLVTRVVTHEMAHA 98
QY 119 -----GGNNAALVNQTASDSSVMVRQVGFNNATANQY 151
DB 99 NNATDQWNAKNSDITVQYGGNNAALVNQTASDSSVMVRQVGFNNATANQY 151
RESULT 9
```

AAB36346
 ID AAB36346 standard; protein; 151 AA.
 AC AAB36346;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Agfa::PT3#1 amino acid sequence SEQ ID NO:12.
 XX
 KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 FN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay WW;
 XX
 DR WPI: 2000-672631/65.
 DR N-PSDB; AAC64622.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 135; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;
 Query Match 78.5%; Score 603; DB 3; Length 151;
 Best Local Similarity 80.8%; Pred. No. 1.7e-50;
 Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;
 Qy 1 MKLLKVAFAAIVVSGSALACVFPQWGGGNGHNGSGPDSTLSIYQYGSANALALQ 60
 DB 1 MKLLKVAFAAIVVSGSALACVFPQWGGGNGHNGSGPDSTLSIYQYGSANALALQ 60

Qy 61 SDARKSETTITQSGYNGADVGQADNSTIELTQGFRRNATYDQLVTRVWTHEVAHAGG 120
 DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQGFRRNATIDQWNAKNSDITVQYGG 120
 Qy 121 NNAALVNQTASDSSVMVROVGFNNATANQY 151
 DB 121 NNAALVNYDQLVTRVWTHEMAHANNATANQY 151

RESULT 10

AAB36353
 ID AAB36353 standard; protein; 151 AA.
 XX
 AC AAB36353;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Agfa::PT3#8 amino acid sequence SEQ ID NO:26.
 XX
 KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 FN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX

PI White AP, Doran JL, Collison SK, Kay WW;
 XX
 DR WPI: 2000-672631/65.
 DR N-PSDB; AAC64629.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 138; 139pp; English.

CC The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in

CC the exemplification of the present invention
XX
SQ Sequence 151 AA;

Query Match 78.1%; Score 600; DB 3; Length 151;
Best Local Similarity 81.5%; Pred. No. 3.4e-50;
Matches 123; Conservative 4; Mismatches 24; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALVDQ 60

QY 61 SDARKSETTITQSGYCGNADVGQADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGG 120
DB 61 LVTRVVTHEMAHAGVGNADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120

QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 11
AAB36349
ID AAB36349 standard; protein; 151 AA.
XX
AC AAB36349;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa::PT3#4 amino acid sequence SEQ ID NO:18.
XX
KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
XX vaccine; immune response; immunogen.
XX
OS Salmonella enteritidis.
XX Escherichia coli.
XX Synthetic.
XX
FN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
XX N-PSDB; AAC64625.
XX

Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 136; 139pp; English.

The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAP) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or

CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell) the hybrid fimbria protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 151 AA;

Query Match 78.0%; Score 599; DB 3; Length 151;
Best Local Similarity 81.5%; Pred. No. 4.2e-50;
Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60

QY 61 SDARKSETTITQSGYCGNADVGQADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGG 120
DB 61 SDARKSETTITQSGYCGNADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120

QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 12
AAB36350
ID AAB36350 standard; protein; 151 AA.
XX
AC AAB36350;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa::PT3#5 amino acid sequence SEQ ID NO:20.
XX
KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
XX vaccine; immune response; immunogen.
XX
OS Salmonella enteritidis.
XX Escherichia coli.
XX Synthetic.
XX
FN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
XX N-PSDB; AAC64626.
XX

Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 137; 139pp; English.

The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAP) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or

CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer or segments of a recombinant AgfA
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX Sequence 151 AA;

Query Match 77.7%; Score 597; DB 3; Length 151;
 Best Local Similarity 81.5%; Pred. No. 6.6e-23;
 Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;
 QY 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGGSGDPSTLSIYQGSANAALALQ 60
 DB 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGGSGDPSTLSIYQGSANAALALQ 60
 QY 61 SDARKSETTITQSGYNGADVQGGADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGG 120
 DB 61 SDARKYDQLVTRVVTHEMAHAGQGGADNSTIELTQNGFRNNATIDOWNAKNSDITVQGYGG 120
 QY 121 NNAALVNOTASDSSVMVQVGFNNATANQY 151
 DB 121 NNAALVNOTASDSSVMVQVGFNNATANQY 151

RESULT 13
 AAB36348
 ID AAB36348 standard; protein; 151 AA.
 AC AAB36348;
 XX 26-FEB-2001 (first entry)
 DE AgfA::PT3#3 amino acid sequence SEQ ID NO:16.
 KW Salmonella; agfA; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX WO2000060102-A2.
 XX 12-OCT-2000.
 PF 05-APR-2000; 2000WO-CA000356.
 XX 05-APR-1999; 99US-0127888P.
 XX (UYVI-) UNIV VICTORIA.
 XX White AP, Doran JL, Collison SK, Kay WW;
 XX WPI; 2000-672631/65.

DR N-PSDB; AAC64624.
 XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant AgfA
 PT protein useful for eliciting immune response in animal.
 XX Disclosures; Page 136; 139pp; English.
 XX The present invention describes a recombinant agfA gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer or segments of a recombinant AgfA
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX Sequence 151 AA;

Query Match 73.7%; Score 566; DB 3; Length 151;
 Best Local Similarity 80.8%; Pred. No. 6.7e-47;
 Matches 122; Conservative 5; Mismatches 24; Indels 0; Gaps 0;
 QY 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGGSGDPSTLSIYQGSANAALALQ 60
 DB 1 MKLLKVAAPAAIIVSGSALAGVVDQLVTRVVTHEMAHAGSGDPSTLSIYQGSANAALALQ 60
 QY 61 SDARKSETTITQSGYNGADVQGGADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGG 120
 DB 61 SDARKSETTITQSGYNGADVQGGADNSTIELTQNGFRNNATIDOWNAKNSDITVQGYGG 120
 QY 121 NNAALVNOTASDSSVMVQVGFNNATANQY 151
 DB 121 NNAALVNOTASDSSVMVQVGFNNATANQY 151

RESULT 14
 AAB36343
 ID AAB36343 standard; protein; 151 AA.
 AC AAB36343;
 XX 26-FEB-2001 (first entry)
 DE Escherichia coli CsgA amino acid sequence SEQ ID NO:7.
 KW Salmonella; agfA; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 OS Escherichia coli.
 XX WO2000060102-A2.
 XX 12-OCT-2000.

PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UUVI-) UNIV VICTORIA.
 XX
 FI White AP, Doran JL, Collison SK, Kay WW;
 XX
 DR WPI; 2000-672631/65.
 DR N-PSDB; AAC64619.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 135; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (Sef1/7fAR) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fibrin subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fibrin protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fibrin subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;
 Query Match 67.3%; Score 517; DB 3; Length 151;
 Best Local Similarity 69.5%; Pred. No. 3.8e-42;
 Matches 105; Conservative 17; Mismatches 29; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
 DB 1 MKLLKVAIAIAIVFSGSALAGVVPQYGGGNGHGGGNSGPNSELNIYQYGGNSALALQ 60
 QY 61 SDARKSETTITQSGYNGADYVCGQADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGG 120
 DB 61 TDARNSDLTITQHGCGGADYVCGQSDSSIDLITQRFNGSATLDQWNGKNSMTVKQFGG 120
 QY 121 NNAALVNQTASDSSVVMVROVGFNNATANQY 151
 DB 121 GNGAAVDQTASNSVNVTVQVFGNNATAHQY 151
 RESULT 15
 ABR82651
 ID ABR82651 standard; protein; 151 AA.
 XX
 AC ABR82651;
 XX
 DT 04-DEC-2003 (first entry)
 XX
 DE E. coli CsgA subunit 15 kDa protein.
 XX

KW Plasma protein; immune response; antibacterial; vaccine; gene therapy.
 XX Escherichia coli.
 XX WO2003064446-A2.
 XX
 PD 07-AUG-2003.
 XX
 XX 30-JAN-2003; 2003WO-EP000943.
 XX
 XX 31-JAN-2002; 2002GB-00002275.
 XX
 XX (HANS-) HANSA MEDICAL RES AB.
 XX
 PI Ejoerck L, Olsen A, Wikstroem M, Herwald H;
 XX
 DR WPI; 2003-646136/61.
 DR N-PSDB; ACF36153.
 XX
 PT New isolated peptide capable of binding a mammalian plasma protein,
 PT useful in the manufacture of a medicament for the prevention and/or
 PT treatment of a bacterial infection, such as Escherichia coli, Salmonella
 PT or Shigella infections.
 XX
 PS Disclosure; Page 41-42; 42pp; English.
 XX
 CC The invention relates to an isolated peptide capable of binding a
 CC mammalian plasma protein or of generating an immune response in a mammal
 CC selected from sequences shown in ABR82642, ABR82648-49. The peptide or
 CC antibody is useful for treating a bacterial infection in a human or
 CC animal or in the manufacture of a medicament for the prophylactic
 CC treatment of a bacterial infection, such as Escherichia coli, Salmonella
 CC or Shigella infection. The peptide that is immobilized on a solid support
 CC is also useful as a reagent for determining the ability of a plasma
 CC protein to bind to bacteria. The present sequence represents an E. coli
 CC 15 kDa protein
 XX
 SQ Sequence 151 AA;
 Query Match 66.7%; Score 512; DB 7; Length 151;
 Best Local Similarity 68.9%; Pred. No. 1.2e-41;
 Matches 104; Conservative 17; Mismatches 30; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
 DB 1 MKLLKVEATAIVFSGSALAGVVPQYGGGNGHGGGNSGPNSELNIYQYGGNSALALQ 60
 QY 61 SDARKSETTITQSGYNGADYVCGQADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGG 120
 DB 61 TDARNSDLTITQHGCGGADYVCGQSDSSIDLITQRFNGSATLDQWNGKNSMTVKQFGG 120
 QY 121 NNAALVNQTASDSSVVMVROVGFNNATANQY 151
 DB 121 GNGAAVDQTASNSVNVTVQVFGNNATAHQY 151
 Search completed: August 2, 2004, 14:48:29
 Job time : 44.9 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:40:48 ; Search time 12 Seconds

(without alignments)
649.627 Million cell updates/sec

Title: US-09-543-407-30

Perfect score: 768

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVGFNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*

2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*

3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*

4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*

5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*

6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	676	88.0	151	1	US-08-233-788A-59
2	496	64.6	120	1	US-08-233-788A-57
3	92.5	12.0	738	3	US-08-864-038A-3
4	91.5	11.9	673	3	US-08-196-387-8
5	91.5	11.9	673	4	US-08-841-835-8
6	91.5	11.9	949	3	US-08-196-387-10
7	91.5	11.9	949	4	US-08-841-835-10
8	91.5	11.9	1327	3	US-08-196-387-2
9	91.5	11.9	1327	4	US-08-841-835-2
10	91.5	11.9	1327	4	US-08-972-115A-8
11	86.5	11.3	745	4	US-08-336-115C-6
12	84	10.9	2736	4	US-08-232-991A-30227
13	83	10.8	1216	4	US-08-134-000C-5130
14	81.5	10.6	589	4	US-08-489-039A-7849
15	80	10.4	867	4	US-08-540-236-2676
16	80	10.4	878	4	US-08-540-236-3401
17	78.5	10.2	309	4	US-08-252-991A-22266
18	78.5	10.2	873	4	US-08-336-447A-13
19	78.5	10.2	892	4	US-08-336-447A-5
20	78.5	10.2	1415	4	US-08-252-991A-26438
21	77.5	10.1	1160	3	US-08-808-599A-24
22	77.5	10.1	1207	4	US-08-489-039A-11518
23	77	10.0	415	4	US-08-025-769B-280
24	77	10.0	975	4	US-08-328-352-4764
25	77	10.0	1003	1	US-08-571-758-4
26	77	10.0	1003	1	US-08-909-984A-4
27	77	10.0	1003	1	US-08-909-983-4

28	77	10.0	1034	4	US-09-252-991A-26658	Sequence 26658, A
29	77	10.0	1890	4	US-09-595-684B-39	Sequence 39, Appl
30	77	10.0	1864	2	US-08-804-227C-3	Sequence 3, Appl
31	77	10.0	2315	4	US-09-543-681A-5434	Sequence 5434, Ap
32	76.5	10.0	624	4	US-09-336-447A-7	Sequence 7, Appl
33	76.5	10.0	941	4	US-09-336-447A-9	Sequence 9, Appl
34	76.5	10.0	1398	1	US-08-750-532-9	Sequence 9, Appl
35	76.5	10.0	1398	3	US-08-894-818B-8	Sequence 8, Appl
36	76.5	10.0	1398	4	US-09-445-472-6	Sequence 6, Appl
37	76.5	10.0	1612	1	US-08-169-927-2	Sequence 2, Appl
38	76.5	10.0	2123	3	US-08-968-685A-10	Sequence 10, Appl
39	76	9.9	231	4	US-09-540-236-3827	Sequence 3827, Ap
40	76	9.9	870	4	US-09-177-650-91	Sequence 91, Appl
41	75.5	9.8	238	4	US-09-495-880A-42	Sequence 42, Appl
42	75.5	9.8	282	2	US-08-860-174A-10	Sequence 10, Appl
43	75.5	9.8	365	4	US-09-489-847-322	Sequence 322, App
44	75.5	9.8	394	4	US-09-252-991A-28148	Sequence 28148, A
45	75.5	9.8	432	4	US-09-403-089A-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-233-788A-59
; Sequence 59, Application US/08233788A
; Patent No.: 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, Karen S.
; APPLICANT: Clouthier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; TITLE OF INVENTION: OF SALMONELLA
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,788A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 59:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 151 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-233-788A-59

Query Match 88.0%; Score 676; DB 1; Length 151;
Best Local Similarity 90.1%; Pred. No. 2, 7e-61;
Matches 136; Conservative 2; Mismatches 13; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPWGGGNGHNGSGDPDSTLSTIYQYGSANALALQ 60

Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGSPDSTLSIYQYGSANAALALQ 60
QY 61 SDARSETTITQSGYNGADYVQGGADNSTIETLQNGFRNNATYDQLVTRVVTHEMAHAGG 120
Db 61 SDARSETTITQSGYNGADYVQGGADNSTIETLQNGFRNNATYDQLVTRVVTHEMAHAGG 120
QY 121 NNAALVNQTASDSSVMVQVGFNNATANQY 151
Db 121 NNPALVNQTASDSSVMVQVGFNNATANQY 151

RESULT 2

US-08-233-788A-57
; Sequence 57, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, Karen S.
; APPLICANT: Clouchier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; TITLE OF INVENTION: OF SALMONELLA
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,788A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERY
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 120 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-233-788A-57

Query Match 64.6%; Score 496; DB 1; Length 120;
Best Local Similarity 87.5%; Pred. No. 4e-43;
Matches 98; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

QY 22 VVPQGGGNGHNGGNSGSPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGADV 81
Db 1 VVPQGGGNGHNGGNSGSPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGADV 60
QY 82 CGGADNSTIETLQNGFRNNATYDQLVTRVVTHEMAHAGGNNALVNQTASDS 133
Db 61 CGGADNSTIETLQNGFRNNATIDQWAKNSDITVQYGGNNAALVNQTASDS 112

RESULT 3

US-08-864-038A-3
; Sequence 3, Application US/08864038A
; Patent No. 6001592
; GENERAL INFORMATION:

; APPLICANT: Kunio NAKASHIMA et al.
; TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE CDNA, VECTOR
; TITLE OF INVENTION: CONTAINING SAID CDNA, HOST CELLS TRANSFORMED WITH SAID
; TITLE OF INVENTION: VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING
; TITLE OF INVENTION: SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBODY
; TITLE OF INVENTION: TO SAID POLYPEPTIDE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: 812-5 Hirano
; STREET: Isshinden
; CITY: Tsu-city
; STATE: Mie-prefecture
; COUNTRY: JAPAN
; ZIP: 514-01
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Microsoft Windows 95
; SOFTWARE: Word Perfect 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/864,038A
; FILING DATE: May 28, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-184459
; FILING DATE: 15-July-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: C. Bruce Hamburg
; REGISTRATION NUMBER: 22,389
; REFERENCE/DOCKET NUMBER: F-5610
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 986-2340
; TELEFAX: (212) 953-7733
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 738
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Pinctada fucata
; CELL TYPE: mantle epithelial cell
; FEATURE:
; NAME/KEY: peptide
; LOCATION: from 1 to 738
; IDENTIFICATION METHOD: E (by experiment)
US-08-864-038A-3

Query Match 12.0%; Score 92.5; DB 3; Length 738;
Best Local Similarity 27.7%; Pred. No. 0.46; Indels 39; Gaps 6;
Matches 43; Conservative 12; Mismatches 61

QY 3 LLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGSPDSTLSIYQYGSANAALALQSD 62
Db 419 LLKSSASASASASASAG-----GGGGGGGGGGGGG-----GGGAGALA----- 460
QY 63 ARKSETTITQSGYNGADVQGGADNSTIETLQNGFRNNATYDQLVTRVVTHEMAHAGGNN 122
Db 461 -----AALAAAGAGGGLGGGGGGGALAAALAAAG-AGGGGPGGL-----GGLGGLGGGS 508
QY 123 AALVNQTASDSS-----VMVQVGFNNATA 148
Db 509 AAAAAAASASASASASAG-----VMVQVGFNNATA 543

RESULT 4

US-09-196-387-8
; Sequence 8, Application US/09196387
; Patent No. 627613
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRP1 AND METHODS
; TITLE OF INVENTION: OF USE THEREOF


```

; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 949 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-196-387-10

Query Match 11.9%; Score 91.5; DB 3; Length 949;
Best Local Similarity 29.4%; Pred. No. 0.8;
Matches 42; Conservative 17; Mismatches 65; Indels 19; Gaps 6;

QY 6 VAAFAAI-VVSGSALAGVVPQWGGGNNHGGNSGDPSTLSIYQYGSANAALALQSDAR 64
DB 99 VAAAPVPAVTSSTSAAGVAPNPAGSGNNSSPSSPTSS-SSSSPSSPGSSLAESPEAA 157
QY 65 KSETTIT----QSGYNGADVQGGADNSTIELTONGFRNNATYDQLVTRVVTHEMAHAGG 120
DB 158 GVSSTAPLPGGAAGPGTGVPAVSGALRELLEACRNGD-----VSRV--KRLVDAAN 206
QY 121 NNAALVNOTASDSSVMVRQVGF 143
DB 207 VNAK--DMAGRKSSPLHPAAGFG 227

RESULT 7
US-09-841-835-10
; Sequence 10, Application US/09841835
; Patent No. 6506587
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; TITLE OF INVENTION: OF USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/841,835
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/196,387
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 949 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

```

```

; MOLECULE TYPE: protein
US-09-841-835-10

Query Match 11.9%; Score 91.5; DB 4; Length 949;
Best Local Similarity 29.4%; Pred. No. 0.8;
Matches 42; Conservative 17; Mismatches 65; Indels 19; Gaps 6;

QY 6 VAAFAAI-VVSGSALAGVVPQWGGGNNHGGNSGDPSTLSIYQYGSANAALALQSDAR 64
DB 99 VAAAPVPAVTSSTSAAGVAPNPAGSGNNSSPSSPTSS-SSSSPSSPGSSLAESPEAA 157
QY 65 KSETTIT----QSGYNGADVQGGADNSTIELTONGFRNNATYDQLVTRVVTHEMAHAGG 120
DB 158 GVSSTAPLPGGAAGPGTGVPAVSGALRELLEACRNGD-----VSRV--KRLVDAAN 206
QY 121 NNAALVNOTASDSSVMVRQVGF 143
DB 207 VNAK--DMAGRKSSPLHPAAGFG 227

RESULT 8
US-09-196-387-2
; Sequence 2, Application US/09196387
; Patent No. 627613
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; TITLE OF INVENTION: OF USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/196,387
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/095,225
; FILING DATE: June 10, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1327 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
US-09-196-387-2

Query Match 11.9%; Score 91.5; DB 3; Length 1327;
Best Local Similarity 29.4%; Pred. No. 1.3;
Matches 42; Conservative 17; Mismatches 65; Indels 19; Gaps 6;

QY 6 VAAFAAI-VVSGSALAGVVPQWGGGNNHGGNSGDPSTLSIYQYGSANAALALQSDAR 64

```

```
Db 99 VAAAPVPAVSTSSAAGVAPNPAAGSGNNSSSSPTSS--SSSPSPGSSLAESPEAA 157
QY 65 KSETTIT----QSGYNGADYQGADNSTIETQNGFNNAATYDQLVTRVVTHEMAHAGG 120
Db 158 GVSSTAPLPGAGGPGTGPVAVSGALRELLACRNGD-----VSRV--KELVDAAN 206
QY 121 NNAALVNQTASDSSVMVRQVGF 143
Db 207 VNAX--DMAGKSSPLHFAAGFG 227

RESULT 9
US-09-841-835-2
; Sequence 2, Application US/09841835
; Patent No. 6506587
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; TITLE OF INVENTION: OF USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/09/841.835
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1327 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
US-09-841-835-2

Query Match
Best Local Similarity 11.9%; Score 91.5; DB 4; Length 1327;
Matches 42; Conservative 17; Mismatches 65; Indels 19; Gaps 6;

QY 6 VAAFAAI-VVSGSALAGVVPQWGGNGNNGSGPDSTLSIYQYGSANAALALQSDAR 64
Db 99 VAAAPVPAVSTSSAAGVAPNPAAGSGNNSSSSPTSS--SSSPSPGSSLAESPEAA 157
QY 65 KSETTIT----QSGYNGADYQGADNSTIETQNGFNNAATYDQLVTRVVTHEMAHAGG 120
Db 158 GVSSTAPLPGAGGPGTGPVAVSGALRELLACRNGD-----VSRV--KELVDAAN 206
QY 121 NNAALVNQTASDSSVMVRQVGF 143
Db 207 VNAX--DMAGKSSPLHFAAGFG 227

RESULT 11
US-09-336-115C-6
; Sequence 6, Application US/09336115C
; Patent No. 6576244
; GENERAL INFORMATION:
; APPLICANT: Weltzin, Richard A.
; APPLICANT: Guy, Bruno
; TITLE OF INVENTION: LT and CT in Parenteral Immunization
; TITLE OF INVENTION: Methods Against Helicobacter Infection
; FILE REFERENCE: 06132/055002
; CURRENT APPLICATION NUMBER: US/09/336,115C
; CURRENT FILING DATE: 1999-06-18
; PRIOR APPLICATION NUMBER: US 09/100,258
; PRIOR FILING DATE: 1998-06-19
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 745
; TYPE: PRT
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)...(20)
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 721
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-336-115C-6
```

```
Query Match      11.3%; Score 86.5; DB 4; Length 745;
Best Local Similarity 20.0%; Pred. No. 1.9;
Matches 34; Conservative 27; Mismatches 72; Indels 37; Gaps 5;

QY 8 AFAAIWGSAGVVPWQ-----GGGNGHNGGN----- 37
DB 102 AVQAVFLAINAAVL---WNTIGYAVMCGNGGTSGPGSVIFNDQPGQDSTQITCNKFE 158
QY 38 SSGPSTLSIYQGSANAALQSDARKSETTITQSGYGCADYQGGADNSTIELTQNGF 97
DB 159 STGPKGNSIDFEKFLNAYQIIQALXQSGFFELG-GNGTKV---SVNYYNECROTAD 214
QY 98 RNNATYDQLVTRVTHEMAHAGNNAALVNQTSDDSSVMVRQVFGNNAT 147
DB 215 INGGVYQCKAKNGSSSSNGGSSSTQTTATTQDGVITITTYNNKAT 264

RESULT 12
US-09-252-991A-30227
; Sequence 30227, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30227
; LENGTH: 2736
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30227

Query Match      10.9%; Score 84; DB 4; Length 2736;
Best Local Similarity 29.7%; Pred. No. 19;
Matches 43; Conservative 17; Mismatches 49; Indels 36; Gaps 8;

QY 6 VAAFAIVVSGSAL-AGVV--PQWGGGNGHNGGNSGPDST-----LSIYQVGSANA 55
DB 149 MSAGTALLVGAAGAGVAINSSGGG---GGGSSVFPDTPPKAASGLKIAPDGSIS 205
QY 56 ALALQSDARKSETTITQSGYGNADV-CQGADNSTIELTQNG-----FRNNATYDQLVTR 109
DB 206 GQA-----EAGASGVIDTNGDKPDLTVIADANGFTAPLNPELTNGQVTV 252
QY 110 VVTHEMAHAGNNAALVNQTSDDSS 134
DB 253 VVT-----DPAGNASPPAQVTAFTTT 273

RESULT 13
US-09-134-000C-5130
; Sequence 5130, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; FILE REFERENCE: 032796-032
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5130
; LENGTH: 1216
```

```
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-5130

Query Match      10.8%; Score 83; DB 4; Length 1216;
Best Local Similarity 24.4%; Pred. No. 8.2;
Matches 49; Conservative 19; Mismatches 55; Indels 78; Gaps 10;

QY 6 VAAFAIVV--SGSALAGVVPWGG--GSHNGGNSGSPDST-----LSIY- 48
DB 181 VTGNSLVNDSSGGRIAS!---YGGYGNTATNTANTVGNSTKVAITNAATGFOLSTYY 237
QY 49 ---QYGSANAALQSDARKSETTITQSGYGNADVQ-----GADNSTIELTQ 94
DB 238 GGQVQGNIG-----GKVTNDISGYGRWYTAGQRFIGGSSRGDIGNRATDGT 286
QY 95 N-----GFRNNATYDQLVTRVTHEMAHAG-----GNNAAALVNQTA 130
DB 287 NLNTQLYSAGRADFEQNGYSGTIIGDITNVVTAGTNSAGGINDFGGAGNNVSKFNK- 345
QY 131 SDSSVMVRQVFGNNATANCY 151
DB 346 -----QIGASNEATYDAY 358

RESULT 14
US-09-489-039A-7849
; Sequence 7849, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 7849
; LENGTH: 589
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-7849

Query Match      10.6%; Score 81.5; DB 4; Length 589;
Best Local Similarity 24.5%; Pred. No. 4.5;
Matches 35; Conservative 21; Mismatches 56; Indels 31; Gaps 7;

QY 29 GGNHNGGNSGSPDSTLSIYQGS-----ANAALQSDARKSETTITQSGYNGA- 79
DB 259 GGLDRNGANANGQTDTFGIYAFDTLTLTIERIEINGGLRLDNYHTKYDSATACGSGRGAI 318
QY 80 --DVQG--GADNSTIELTONG-----FRNNATYDQLVTRVTHEMAHAGN---NAALVNQ 128
DB 319 ACPFGQSTGSPVTVDTAKSGLNVNWKAGALYR-----LTBQ-----GNVYVNTAISQ 367
QY 129 TASDSSVMVRQVFGNNATANCY 151
DB 368 PGGSSFALAAASGSGNSANRTDF 390

RESULT 15
US-09-540-236-2676
; Sequence 2676, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAR
; FILE REFERENCE: 2709.2005-001
; CURRENT FILING DATE: 2000-04-04
```


This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:54:48 ; Search time 36.8 Seconds
(without alignments)
1287.123 Million cell updates/sec

Title: US-09-543-407-30
Perfect score: 768
Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVFGNNATANYQ 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	514	66.9	151	12	US-09-741-873B-4
2	514	66.9	151	12	US-09-741-873B-4
3	436	56.8	131	12	US-09-741-873B-2
4	436	56.8	131	12	US-09-741-873B-2
5	103	13.4	445	15	US-10-369-493-20638
6	94.5	12.3	537	16	US-10-437-963-141342
7	94	12.2	354	10	US-09-820-843A-21
8	94	12.2	498	12	US-10-424-599-196154
9	93.5	12.2	2204	12	US-10-282-122A-64364
10	93	12.1	263	12	US-10-425-114-49860
11	92.5	12.0	1721	12	US-10-282-122A-62548
12	91.5	11.9	673	9	US-09-841-835-8
13	91.5	11.9	949	9	US-09-841-835-10
14	91.5	11.9	1327	9	US-09-841-835-2
15	91.5	11.9	1327	10	US-09-972-115A-8

16	91.5	11.9	1327	14	US-10-199-937-4	Sequence 4, Appli
17	90.5	11.8	628	12	US-10-282-122A-53269	Sequence 53269, A
18	90	11.7	171	12	US-10-424-599-173860	Sequence 173860,
19	90	11.7	703	16	US-10-437-963-108981	Sequence 108981, A
20	89.5	11.7	1129	12	US-10-282-122A-48048	Sequence 48048, A
21	88.5	11.5	145	16	US-10-437-963-147748	Sequence 147748,
22	88.5	11.5	1217	14	US-10-311-406-2	Sequence 2, Appli
23	88.5	11.5	1862	12	US-10-282-122A-49757	Sequence 49757, A
24	87.5	11.4	191	15	US-10-437-963-105413	Sequence 105413, A
25	87	11.3	369	12	US-10-425-114-56041	Sequence 56041, A
26	87	11.3	486	12	US-10-424-599-275468	Sequence 275468, A
27	87	11.3	507	12	US-10-425-114-57763	Sequence 57763, A
28	87	11.3	597	9	US-09-793-306-146	Sequence 146, App
29	86.5	11.3	197	12	US-10-425-114-67750	Sequence 67750, A
30	86.5	11.3	745	8	US-08-834-666A-6	Sequence 6, Appli
31	86.5	11.3	745	12	US-10-282-122A-58683	Sequence 58683, A
32	86	11.2	538	12	US-10-425-114-68152	Sequence 68152, A
33	85.5	11.1	270	16	US-10-437-963-122263	Sequence 122263, A
34	85.5	11.1	298	12	US-10-282-122A-47486	Sequence 47486, A
35	85	11.1	321	15	US-10-412-699B-1074	Sequence 1074, App
36	85	11.1	321	15	US-10-374-780A-590	Sequence 590, App
37	85	11.1	321	15	US-10-437-963-172276	Sequence 172276,
38	85	11.1	323	12	US-10-659-869-26	Sequence 26, Appl
39	85	11.1	323	14	US-10-021-811-26	Sequence 26, Appl
40	85	11.1	1778	14	US-10-238-075-749	Sequence 749, App
41	84.5	11.0	322	9	US-09-905-176-24	Sequence 24, Appl
42	84.5	11.0	678	12	US-10-282-122A-64573	Sequence 64573, A
43	84	10.9	65	9	US-09-996-194-16	Sequence 16, Appl
44	84	10.9	65	12	US-10-164-966-33	Sequence 33, Appl
45	84	10.9	478	16	US-10-437-963-115033	Sequence 115033,

ALIGNMENTS

RESULT 1

US-09-741-873B-4
; Sequence 4; Application US/09741873B
; Publication No. US20020081722A1
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; PRIOR FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-4

Query Match 66.9%; Score 514; DB 12; Length 151;
Best Local Similarity 68.9%; Pred. NO. 1.4e-43;
Matches 104; Conservative 18; Mismatches 29; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAVVVPQWGGGNGGNSGPDSTLSIYQYGSANAALQ 60

[illegible]

RESULT 2
US-09-741-873B-4
; Sequence 4, Application US/09741873B
; Publication No. US20040096965A9
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsner, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B

```

/ CURRENT FILING DATE: 2003-04-04
/ PRIOR APPLICATION NUMBER: SE 8801723-1
/ PRIOR FILING DATE: 1998-05-06
/ PRIOR APPLICATION NUMBER: US 08/978,878
/ PRIOR FILING DATE: 1997-11-26
/ PRIOR APPLICATION NUMBER: US 07/347,189
/ PRIOR FILING DATE: 1989-05-04
/ PRIOR APPLICATION NUMBER: US 07/789,437
/ PRIOR FILING DATE: 1991-11-06
/ PRIOR APPLICATION NUMBER: US 07/970,846
/ PRIOR FILING DATE: 1992-11-03
/ PRIOR APPLICATION NUMBER: US 08/187,865
/ PRIOR FILING DATE: 1994-01-28
/ PRIOR APPLICATION NUMBER: US 08/318,519
/ PRIOR FILING DATE: 1994-10-05
/ NUMBER OF SEQ ID NOS: 10
/ SOFTWARE: PatentIn version 3.0
/ SEQ ID NO 4
/ LENGTH: 151
/ TYPE: PR1
/ ORGANISM: Escherichia coli
/ US-09-741-871R-4

```

```

RESULT 3
US-09-741-873B-2
; Sequence 2, Application US/09741873B
; Publication NO. US20020081722A1
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olser, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012869-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; CURRENT FILING DATE: 2003-04-04

```

```
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-2

Query Match      56.8%; Score 436; DB 12; Length 131;
Best Local Similarity 65.6%; Pred. No. 7.7e-35;
Matches 86; Conservative 17; Mismatches 28; Indels 0; Gaps 0

Qy    21  GVVPQWGGGNGHNGGNSGPDLSIYQYGSAANAALOSDARKSETTITQSGYGNGAD 80
       :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db     1  GVVPQYGGGNGHGCGNNSGENSELNTYQYGGGSALALQTDARNSDLTIHQGGGNGAD 60
       :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Qy    81  VCGGDNSTIELTQGFENNAATYPOLTVRVVTHEMAHGNNALVNQTASDSVVNVROV 140
       :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db     61  VCQGDDSDSIDLTQRGFGNSALTLDQWGNKNSMTVKQFGGGNGAAVDQTASNSSVNVTVQ 120
       :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Qy    141  GFGNNTANQY 151
       |||||::|::|
Db     121  GFGNNTAHOY 131
       |||||::|::|
```

```

RESULT 4
US-09-741-873B-2
; Sequence 2, Application US/09741873B
; Publication No. US20040096965A9
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; CURRENT FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-2
Query Match          56.8%;   Score 436;   DB 12;   Length 131;

```



```
Best Local Similarity 65.6%; Pred. No. 7.7e-36;
Matches 86; Conservative 17; Mismatches 28; Indels 0; Gaps 0;

Qy 21 GVVPQGGGNGHNGGSSGPDSTLSIYQGSANAALALQSDARKSETTITQSGYNGAD 80
Db 1 GVVPQGGGNGHNGGSSGPNSEINTYQYGGNSALALQTDARNSDLTIHQGGNGAD 60

Qy 81 VQGGADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGNNAALVNQTSASSVVMVROV 140
Db 61 VQGGSDSSIDLTFQCGFNSATLDQWNGKNSMTVKQFGGNGAAVDQTSASSVNVTOV 120

Qy 141 GFGNNATANQY 151
Db 121 GFGNNATAHOY 131

RESULT 5
US-10-369-493-20638
; Sequence 20638, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 20638
; LENGTH: 445
; TYPE: PRT
; ORGANISM: Rhodospseudomonas palustris
; NAME/KEY: unsure
; FEATURE:
; LOCATION: (1)..(445)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-20638

Query Match 13.4%; Score 103; DB 15; Length 445;
Best Local Similarity 26.3%; Pred. No. 0.079;
Matches 44; Conservative 27; Mismatches 62; Indels 34; Gaps 8;

Qy 7 AAFAA-----IVVSGSALAGVFPQWGGG-----NHNNG-----GNSSGPDSTLSIYQY 50
Db 19 AAFRAADGNTVYLNQTDNQANITQSGNGSVGAFNGSGFLQENGLTSGA-NLLTVKQS 77

Qy 51 GSANAALALQSDARKSETTITQSGYNGADYVQGGADNSTIELTQNGFRNNA-----TY 103
Db 78 GNSNSV-----GRDIQGGQSGAGNSAALFQEGTQSDVELQQTGTSGNAVPSGWNWTN 129

Qy 104 DQLVTRVVTHEMAHAGNNAALVNQTSASSVVMVROVQFGNNATANQ 150
Db 130 DPGVFNKTIQD-SSNSGKSVI-QDGKNVFSIKQGTNGTNGTSVQ 174

RESULT 6
US-10-437-963-141342
; Sequence 141342, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
```

```
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 141342
; LENGTH: 537
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_42455C.1.pap
US-10-437-963-141342

Query Match 12.3%; Score 94.5; DB 16; Length 537;
Best Local Similarity 27.1%; Pred. No. 0.71;
Matches 45; Conservative 17; Mismatches 55; Indels 49; Gaps 9;

Qy 3 LLKVAAPAAIIVVSGSALAGVFPQWGGG-----GHNHNGG-----NNS 39
Db 347 LLNLGFF-----SGS--NGGYDQFNGGAGNGGGSIVTSSGLAGNHGGGGFSLYNSS 399

Qy 40 QPDSTLSIYQGSANAALALQSDARKSETTITQS--GYNGADYVQGGADNSTIELTQNGF 97
Db 400 EPAGTLP-----QMSATALLQAAQMGATTSSYNAGGAGGASSLLRGASSHGISVGECPA 454

Qy 98 RNNATYDQLVTRVVTHEMAHAGNNAALVNQTSASSVVMVROVQVGF 143
Db 455 NERSYQNL-----IMGSA-SGGGAGPAGFSGAS-----GEG 488

RESULT 7
US-09-820-843A-21
; Sequence 21, Application US/09820843A
; Publication No. US20030039963A1
; GENERAL INFORMATION:
; APPLICANT: Council of Scientific and Industrial Research
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTE
; FILE REFERENCE: 063915
; CURRENT APPLICATION NUMBER: US/09/820,843A
; CURRENT FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 21
; LENGTH: 354
; TYPE: PRT
; ORGANISM: M. tuberculosis
; NAME/KEY: misc.feature
; OTHER INFORMATION: PPE
; NAME/KEY: misc.feature
; OTHER INFORMATION: gi|1781260
US-09-820-843A-21

Query Match 12.2%; Score 94; DB 10; Length 354;
Best Local Similarity 25.0%; Pred. No. 0.47;
Matches 35; Conservative 16; Mismatches 71; Indels 18; Gaps 5;

Qy 26 WGGGNGHNGGNSGSPDS---TLSIYQGSANAALALQSDARKSETTI-----TQSGY 75
Db 148 WNGGSTNTGLANAGAGNTGFFDAGNYNFGSLNAGNINSSFGNSGDNGLNAGDVNSGV 207

Qy 76 GNGADYVQGGADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAG-----GNNALVNQT 129
Db 208 GNAGDVNTGLNSG-NINTGGF-NPGLTNTGFFSANTQAGPNSGFFNAGTSGNSGFCHNDP 265

Qy 130 ASDSSVMVROVQFGNNATAN 149
Db 266 AGSGNSGIGNSGFGNSGYVN 285
```

RESULT 8
 US-10-424-599-196154
 ; Sequence 196154, Application US/10424599
 ; Publication No. US20040031072A1
 ; GENERAL INFORMATION:
 ; APPLICANT: La Rosa Thomas J
 ; APPLICANT: Kovalic David K
 ; APPLICANT: Zhou Yihua
 ; APPLICANT: Cao Yongwei
 ; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
 ; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 ; FILE REFERENCE: 38-21(53223)B
 ; CURRENT APPLICATION NUMBER: US/10/424,599
 ; CURRENT FILING DATE: 2003-04-28
 ; NUMBER OF SEQ ID NOS: 285684
 ; SEQ ID NO 196154
 ; LENGTH: 498
 ; TYPE: PRT
 ; ORGANISM: Glycine max
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: PAT_MRT3847_19152C.1.pap
 US-10-424-599-196154

Query Match 12.2%; Score 94; DB 12; Length 498;
 Best Local Similarity 28.1%; Pred. No. 0.72;
 Matches 34; Conservative 24; Mismatches 41; Indels 22; Gaps 6;
 QY 30 GNHGGNNGSGPD-STLSIYQYGSANAALALQSDARKSETTITQSGYNG-----A 79
 Db 355 GSSNDNNGSDDTSSEONKEGSDNS--NDANONGNSNENANDNGNAGHEAONNA 412
 QY 80 DVGGADNSTIEL-----TONGFRNNATYDQVTRVWTHEMAHAGG--NNAALVNQTSADSS 134
 Db 413 DAGQGNNEATVESKASENEGGAQNETVESQ-----KEESAHSNGSDNSNLNQGGSDHS 467
 QY 135 V 135
 Db 468 I 468

RESULT 9
 US-10-282-122A-64364
 ; Sequence 64364, Application US/10282122A
 ; Publication No. US20040029129A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Liangsu
 ; APPLICANT: Zamudio, Carlos
 ; APPLICANT: Malone, Cheryl
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari
 ; APPLICANT: Zyskind, Judith
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John
 ; APPLICANT: Carr, Grant
 ; APPLICANT: Yamamoto, Robert
 ; APPLICANT: Forsyth, R.
 ; APPLICANT: Xu, H.
 ; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
 ; FILE REFERENCE: ELITRA.034A
 ; CURRENT APPLICATION NUMBER: US/10/282,122A
 ; CURRENT FILING DATE: 2003-02-20
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/230,335
 ; PRIOR FILING DATE: 2000-09-06
 ; PRIOR APPLICATION NUMBER: 60/230,347
 ; PRIOR FILING DATE: 2000-09-09
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/267,636
 ; PRIOR FILING DATE: 2001-02-09
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 78614
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 64364
 ; LENGTH: 2204
 ; TYPE: PRT
 ; ORGANISM: Mycobacterium tuberculosis
 US-10-282-122A-64364
 Query Match 12.2%; Score 93.5; DB 12; Length 2204;
 Best Local Similarity 26.1%; Pred. No. 5.1;
 Matches 43; Conservative 18; Mismatches 63; Indels 41; Gaps 6;
 QY 10 AAIVVSG-----SALAGV-----VPWGGGNGHNGGSSGPDSTLSIYQYGS 53
 Db 840 AAITVGGPTTSIGITASAGISITPIIDIPATSGFN-----STTSPSSGFFNNGSAGSA 894
 QY 54 NAALALQSDARKSETTITQSGYNGAGDVGGADN-----STIELTQNGFRNNAT 102
 Db 895 SGFLNVVACAGSI-----SGVLNVGALGSGVTNVGHTVGSFYNASALDLVTPAFASGLM 948
 QY 103 YDQVTRVWTHEMAHAGGNAALVNQTSADSSVM---VRQVGFNG 144
 Db 949 RDGMGTMTLNLGLANLGNAGFGNTGTFDVGVANLGNYNIGFGN 993
 RESULT 10
 US-10-425-114-49960
 ; Sequence 49960, Application US/10425114
 ; Publication No. US20040034888A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Liu, Jingdong
 ; APPLICANT: Zhou, Yihua
 ; APPLICANT: Kovalic, David K.
 ; APPLICANT: Screen, Steven E
 ; APPLICANT: Tabaska, Jack E
 ; APPLICANT: Cao, Yongwei
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
 ; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 ; FILE REFERENCE: 38-21(53313)B
 ; CURRENT APPLICATION NUMBER: US/10/425,114
 ; CURRENT FILING DATE: 2003-04-28
 ; NUMBER OF SEQ ID NOS: 73128
 ; SEQ ID NO 49960
 ; LENGTH: 263
 ; TYPE: PRT
 ; ORGANISM: Zea mays
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: 700071884_FLI.pap
 US-10-425-114-49960
 Query Match 12.1%; Score 93; DB 12; Length 263;
 Best Local Similarity 26.8%; Pred. No. 0.41;
 Matches 33; Conservative 16; Mismatches 42; Indels 32; Gaps 5;
 QY 17 SALAGVVPWGGGNGHNGGSSGPDSTLSIYQYGSANAALALQSDARKSETTITQSG-- 74
 Db 84 SSIAG-----GGGGGQGGGCTNGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGG 138
 QY 75 -----YNGGA--DVGGADNSTIEL--TQNGFRNNATYDQVTRVWTHEMAHAGG 120
 Db 139 GGMGGGANGAYGSGAGGCGVGKGVGSGVALAPSSNGYNGCAAD-----ATGGG 187
 QY 121 NNA 123
 ;

Db 188 SGA 190

RESULT 11

US-10-282-122A-62548

Sequence 62548, Application US/10282122A

Publication No. US20040029129A1

GENERAL INFORMATION:

APPLICANT: Wang, Liangsu

APPLICANT: Zamudio, Carlos

APPLICANT: Malone, Cheryl

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Karl

APPLICANT: Zyskind, Judith

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John

APPLICANT: Carr, Grant

APPLICANT: Yamamoto, Robert

APPLICANT: Forsyth, R.

APPLICANT: Xu, H.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

FILE REFERENCE: ELITRA.034A

CURRENT APPLICATION NUMBER: US/10/282,122A

CURRENT FILING DATE: 2003-02-20

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/230,335

PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: 60/230,347

PRIOR FILING DATE: 2000-09-09

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/267,636

PRIOR FILING DATE: 2001-02-09

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 78614

SOFTWARE: PatentIn version 3.1

SEQ ID NO 62548

LENGTH: 1721

TYPE: PRT

ORGANISM: Mycobacterium bovis

US-10-282-122A-62548

Query Match 12.0%; Score 92.5; DB 12; Length 1721;

Best Local Similarity 25.5%; Pred. No. 4.8;

Matches 39; Conservative 18; Mismatches 58; Indels 17; Gaps 6;

Qy 23 VPQWGGGNGHGGNS-----SGPDSTLSIYQY-GSANAALALQSDAKSETTITQSGYGN 77

Db 921 IPATPGFGNSTGGUSSGFFNSGAGSGFGNFGAAGFNNLVS-----TTSGMSGFLN 974

Qy 78 GADYGGQADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGGNNAALVNQTSADSSVMV 137

Db 975 VGALGSGVAN--VGNITSIGIYVGTSD-LSTPAVNSGLANIGTNIAGLLRDGAGTAAI-- 1029

Qy 138 RQVGFNNATAN 149

Db 1030 -NLGLAHGNLN 1040

RESULT 12

US-09-841-835-8

Sequence 8, Application US/09841835

Patent No. US20020076795A1

GENERAL INFORMATION:

APPLICANT: de Lange, Titia

APPLICANT: Smith, Susan

TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSEE: Klauber & Jackson

STREET: 411 Hackensack Avenue, 4th Floor

CITY: Hackensack

STATE: New Jersey

COUNTRY: USA

ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA: US/09/841,835

APPLICATION NUMBER: 09/196,387

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

ATTORNEY/AGENT INFORMATION:

NAME: Jackson Esq., David A.

REGISTRATION NUMBER: 26,742

REFERENCE/DOCKET NUMBER: 600-1-230 CIPI

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201-487-5800

TELEFAX: 201-343-1684

TELEX: 133521

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 673 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-841-835-8

Query Match 11.9%; Score 91.5; DB 9; Length 673;

Best Local Similarity 29.4%; Pred. No. 1.9;

Matches 42; Conservative 17; Mismatches 65; Indels 19; Gaps 6;

Qy 6 VAAFAAI-VVSGSALAGVVPQWGGGNGHGGNSGPDSTLSIYQYGSANAALALQSDAR 64

Db 99 VAAFPVVPVAVSTSSAAGVAPNPAGSGNSPSSSSPTSS-SSSSPSPGSGSLAESPEAA 157

Qy 65 KSETTIT---QSGYGNADYVQGGADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGG 120

Db 158 GVSSTAPLPGGAAGPGTGVPVAVSGALRELLACRNGD-----VSRV--KRLVDAAN 206

Qy 121 NNAALVNQTSADSSVMVQVGF 143

Db 207 VNAAK--DMAGRKSPHPAAGFG 227

RESULT 13

US-09-841-835-10

Sequence 10, Application US/09841835

Patent No. US20020076795A1

GENERAL INFORMATION:

APPLICANT: de Lange, Titia

APPLICANT: Smith, Susan

TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSEE: Klauber & Jackson

STREET: 411 Hackensack Avenue, 4th Floor

CITY: Hackensack

```

; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/841,835
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/196,387
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 949 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-841-835-10

Query Match 11.9%; Score 91.5; DB 9; Length 949;
Best Local Similarity 29.4%; Pred. No. 2.9;
Matches 42; Conservative 17; Mismatches 65; Indels 19; Gaps 6;

QY 6 VAAFAAI-VVSGSALAGVVPWGGGNGHNGGSGDPSTLSIYQYGSANAALALQSDAR 64
Db 99 VAAAPVVPVAVSTSSAAGVAPNPAGSGNNPSSSSPTSS-SSSSPSPGSSLAESPEAA 157

QY 65 KSETTIT---QSGYNGADVGQGDNSTIELTQNGFRNNATYDQLVTRVVVTHEMAHAGG 120
Db 158 GVSSTAPLPGGAAGGTGVPVAVSGALRELEACRNGD-----VSRV--KRLVDAAN 206

QY 121 NNAALVNQTASDSSVMVRQVGF 143
Db 207 VNAX--DMAGRKSSPLHFAAGFG 227

RESULT 14
US-09-841-835-2
; Sequence 2, Application US/09841835
; Patent No. US20020076795A1
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; TITLE OF INVENTION: OF USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/841,835
; FILING DATE:

```

```

; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/196,387
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1327 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; US-09-841-835-2

Query Match 11.9%; Score 91.5; DB 9; Length 1327;
Best Local Similarity 29.4%; Pred. No. 4.3;
Matches 42; Conservative 17; Mismatches 65; Indels 19; Gaps 6;

QY 6 VAAFAAI-VVSGSALAGVVPWGGGNGHNGGSGDPSTLSIYQYGSANAALALQSDAR 64
Db 99 VAAAPVVPVAVSTSSAAGVAPNPAGSGNNPSSSSPTSS-SSSSPSPGSSLAESPEAA 157

QY 65 KSETTIT---QSGYNGADVGQGDNSTIELTQNGFRNNATYDQLVTRVVVTHEMAHAGG 120
Db 158 GVSSTAPLPGGAAGGTGVPVAVSGALRELEACRNGD-----VSRV--KRLVDAAN 206

QY 121 NNAALVNQTASDSSVMVRQVGF 143
Db 207 VNAX--DMAGRKSSPLHFAAGFG 227

RESULT 15
US-09-972-115A-8
; Sequence 8, Application US/09972115A
; Publication No. US20030032769A1
; GENERAL INFORMATION:
; APPLICANT: Geron Corporation
; APPLICANT: Walter, Funk B.
; APPLICANT: Mieczyslaw, Piatyszek A.
; TITLE OF INVENTION: A Second Mammalian Telomerase
; FILE REFERENCE: 080/003C
; CURRENT APPLICATION NUMBER: US/09/972,115A
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: US 60/128,577
; PRIOR FILING DATE: 2000-04-10
; PRIOR APPLICATION NUMBER: US 60/129,123
; PRIOR FILING DATE: 1999-04-13
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: Patent In version 3.1
; SEQ ID NO 8
; LENGTH: 1327
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-972-115A-8

Query Match 11.9%; Score 91.5; DB 10; Length 1327;
Best Local Similarity 29.4%; Pred. No. 4.3;
Matches 42; Conservative 17; Mismatches 65; Indels 19; Gaps 6;

QY 6 VAAFAAI-VVSGSALAGVVPWGGGNGHNGGSGDPSTLSIYQYGSANAALALQSDAR 64
Db 99 VAAAPVVPVAVSTSSAAGVAPNPAGSGNNPSSSSPTSS-SSSSPSPGSSLAESPEAA 157

QY 65 KSETTIT---QSGYNGADVGQGDNSTIELTQNGFRNNATYDQLVTRVVVTHEMAHAGG 120

```

Db 158 GVSSTAPLPGAGPGTGVPAVSGALRELLLEACKNGD-----VSRV--KRLVDAAN 206
Qy 121 NNAALVNQTASDSSVMVRQVGFG 143
Db 207 VNAK--DVAGRKSSFLHFAAGFG 227

Search completed: August 2, 2004, 15:36:14
Job time : 37.8 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:48:33 ; Search time 167.9 Seconds
(without alignments)
877.809 Million cell updates/sec

Title: US-09-543-407-30
Perfect score: 768
Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVFGNNATANY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 6019581 seqs, 976053577 residues

Total number of hits satisfying chosen parameters: 6019581

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents AA Main.*
1: /cgn2_6/ptodata/2/paa/PCTUS COMB.pcp.*
2: /cgn2_6/ptodata/2/paa/US06 COMB.pcp.*
3: /cgn2_6/ptodata/2/paa/US07 COMB.pcp.*
4: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
5: /cgn2_6/ptodata/2/paa/US081 COMB.pcp.*
6: /cgn2_6/ptodata/2/paa/US082 COMB.pcp.*
7: /cgn2_6/ptodata/2/paa/US083 COMB.pcp.*
8: /cgn2_6/ptodata/2/paa/US084 COMB.pcp.*
9: /cgn2_6/ptodata/2/paa/US085 COMB.pcp.*
10: /cgn2_6/ptodata/2/paa/US086 COMB.pcp.*
11: /cgn2_6/ptodata/2/paa/US087 COMB.pcp.*
12: /cgn2_6/ptodata/2/paa/US088 COMB.pcp.*
13: /cgn2_6/ptodata/2/paa/US089 COMB.pcp.*
14: /cgn2_6/ptodata/2/paa/US090 COMB.pcp.*
15: /cgn2_6/ptodata/2/paa/US091 COMB.pcp.*
16: /cgn2_6/ptodata/2/paa/US092 COMB.pcp.*
17: /cgn2_6/ptodata/2/paa/US093 COMB.pcp.*
18: /cgn2_6/ptodata/2/paa/US094 COMB.pcp.*
19: /cgn2_6/ptodata/2/paa/US095 COMB.pcp.*
20: /cgn2_6/ptodata/2/paa/US096 COMB.pcp.*
21: /cgn2_6/ptodata/2/paa/US097A COMB.pcp.*
22: /cgn2_6/ptodata/2/paa/US097B COMB.pcp.*
23: /cgn2_6/ptodata/2/paa/US098 COMB.pcp.*
24: /cgn2_6/ptodata/2/paa/US099A COMB.pcp.*
25: /cgn2_6/ptodata/2/paa/US099B COMB.pcp.*
26: /cgn2_6/ptodata/2/paa/US100 COMB.pcp.*
27: /cgn2_6/ptodata/2/paa/US101 COMB.pcp.*
28: /cgn2_6/ptodata/2/paa/US102 COMB.pcp.*
29: /cgn2_6/ptodata/2/paa/US103 COMB.pcp.*
30: /cgn2_6/ptodata/2/paa/US104 COMB.pcp.*
31: /cgn2_6/ptodata/2/paa/US106 COMB.pcp.*
32: /cgn2_6/ptodata/2/paa/US107 COMB.pcp.*
33: /cgn2_6/ptodata/2/paa/US60 COMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description

1	768	100.0	151	19	US-09-543-407-30	Sequence 30, Appl
2	700	91.1	151	19	US-09-543-407-24	Sequence 24, Appl
3	681	88.7	151	19	US-09-543-407-5	Sequence 5, Appl
4	676	88.0	151	6	US-08-233-642A-57	Sequence 57, Appl
5	659	85.8	151	19	US-09-543-407-14	Sequence 14, Appl
6	657	85.5	151	19	US-09-543-407-22	Sequence 22, Appl
7	613	78.8	151	19	US-09-543-407-28	Sequence 28, Appl
8	603	78.5	151	19	US-09-543-407-12	Sequence 12, Appl
9	600	78.1	151	19	US-09-543-407-26	Sequence 26, Appl
10	599	78.0	151	19	US-09-543-407-18	Sequence 18, Appl
11	597	77.7	151	19	US-09-543-407-20	Sequence 20, Appl
12	594	77.3	131	19	US-09-543-407-31	Sequence 31, Appl
13	566	73.7	151	19	US-09-543-407-16	Sequence 16, Appl
14	517	67.3	151	19	US-09-543-407-7	Sequence 7, Appl
15	514	66.9	151	13	US-08-978-878-4	Sequence 4, Appl
16	514	66.9	151	21	US-09-741-873B-4	Sequence 4, Appl
17	512	66.7	151	33	US-60-352-946-2	Sequence 2, Appl
18	512	66.7	151	33	US-60-444-371-2	Sequence 2, Appl
19	496	64.6	120	6	US-08-233-642A-55	Sequence 55, Appl
20	475	61.8	158	16	US-09-252-691-5834	Sequence 5834, Ap
21	475	61.8	158	16	US-09-252-691C-5834	Sequence 5834, Ap
22	475	61.8	158	30	US-10-417-886-5834	Sequence 34, Appl
23	462	60.2	109	19	US-09-543-407-34	Sequence 2, Appl
24	436	56.8	131	13	US-08-978-878-2	Sequence 35, Appl
25	436	56.8	131	21	US-08-741-873B-2	Sequence 37, Appl
26	332	43.2	109	19	US-09-543-407-35	Sequence 39, Appl
27	310	40.4	68	19	US-09-543-407-37	Sequence 32, Appl
28	237	30.9	48	19	US-09-543-407-39	Sequence 5833, Ap
29	217.5	28.3	70	19	US-09-543-407-32	Sequence 5833, Ap
30	114.5	14.9	186	16	US-09-252-691-5833	Sequence 5833, Ap
31	114.5	14.9	186	16	US-09-252-691C-5833	Sequence 8, Appl
32	114.5	14.9	186	30	US-10-417-886-5833	Sequence 20638, A
33	104.5	13.6	151	19	US-09-543-407-8	Sequence 20638, A
34	103	13.4	445	29	US-10-363-493-20638	Sequence 21, Appl
35	103	13.4	445	33	US-60-360-039-20638	Sequence 6, Appl
36	101.5	13.2	520	1	PCT-US002-18256-21	Sequence 8854, Ap
37	99.5	13.0	151	19	US-09-543-407-6	Sequence 8854, Ap
38	98	12.8	145	21	US-09-733-449-8854	Sequence 22325, A
39	98	12.8	145	23	US-09-803-110-8854	Sequence 22325, A
40	98	12.8	375	21	US-09-733-089-22325	Sequence 22325, A
41	98	12.8	375	23	US-09-816-660-22325	Sequence 23333, A
42	98	12.8	563	21	US-09-733-089-23333	Sequence 23333, A
43	98	12.8	563	23	US-09-816-660-23333	Sequence 9129, Ap
44	96.5	12.6	313	21	US-09-708-427-9129	Sequence 359, App
45	96.5	12.6	1249	30	US-10-455-719-358	

ALIGNMENTS

RESULT 1
US-09-543-407-30
Sequence 30, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIBRILLAR SYSTEM FOR PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543,407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 30
LENGTH: 151
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.

US-09-543-407-30

Query Match	100.0%;	Score 768;	DB 19;	Length 151;
Best Local Similarity	100.0%;	Pred. No. 7e-73;		
Matches 151;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
1	MKLLKVAAPAAALV	SGSALAGV	VPWGCGGHHNGGSSG	PDSTLSIYQYGSANAALQ 60
1	MKLLKVAAPAAALV	SGSALAGV	VPWGCGGHHNGGSSG	PDSTLSIYQYGSANAALQ 60
61	SDARKSETTIT	OSGYCGNADV	QGQADNSTIELT	QTQGFENNATYDQLVTRVVTTHMAHAGG 120
61	SDARKSETTIT	OSGYCGNADV	QGQADNSTIELT	QTQGFENNATYDQLVTRVVTTHMAHAGG 120
121	NNAALVNQTASDSS	VWVRQVGF	GNNATANQY 151	
121	NNAALVNQTASDSS	VWVRQVGF	GNNATANQY 151	

Query Match	91.1%;	Score 700;	DB 19;	Length 151;
Best Local Similarity	89.9%;	Pred. No. 1.2e-65;		
Matches 143;	Conservative 0;	Mismatches 0;	Indels 16;	Gaps 2;

QY	1	MKLLKVAAPAAIIVVSGSALAGVVPQVGCGGNHNGGNSGPDSTLSIYQVGSNAALAQ	60
Db	1	MKLLKVAAPAAIIVVSGSALAGVVPQVGCGGNHNGGNSGPDSTLSIYQVGSNAALAQ	60
QY	61	SPARKSETTITQSGYNGADVGCGADNSITELTQNGFRNNAT	112
Db	61	SPARKSETTITQSGYNGADVGCGADNSITELTQNGFRNNATIDQNAKYDQLVTRVVT	120
QY	113	HEMAHAGGNAAALVNOTASDSSVVMVQVFGNNATANQY	151
Db	121	HEMAHA-----NOTASDSSVVMVQVFGNNATANQY	151

```

; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 151
; TYPE: Prt
; ORGANISM: Salmonella enteritidis
US-09-543-407-5

```

Query Match	88.7%;	Score 681;	DB 19;	Length 151;
Best Local Similarity	90.7%;	Pred. No. 1.2e-63;		
Matches 137;	Conservative 2;	Mismatches 12;	Indels 0;	Gaps 0;
Qy	1	MKLLKVAFAAATVSGSALAGVVPQWGGGNGHGGNSGPDSTLSIYGYGSANAALAQ	60	
Db	1	MKLLKVAFAAATVSGSALAGVVPQWGGGNGHGGNSGPDSTLSIYGYGSANAALAQ	60	
Qy	61	SDARKSETTITQSGYCGADVCGQADNSTIELTQNGFRNATYDQLVTRVVTHEMAHAGG	120	
Db	61	SDARKSETTITQSGYCGADVCGQADNSTIELTQNGFRNATIDQWNAKNSDITVGYGG	120	
Qy	121	NNAALVNQTPADSSVMVRQVFGNNATANQY	151	
Db	121	NNAALVNQTPADSSVMVRQVFGNNATANQY	151	

Query Match	88.0%;	Score	67%;	DB	6;	Length	151;
Best Local Similarity	90.1%;	Pred. NO.	4.28-63;				
Matches	136;	Conservative	2;	Mismatches	13;	Indels	0;
				Gaps	0;		


```

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGSSGDPSTLSIYQYGSANAALALQ 60
Db 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGSSGDPSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVQGGADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGG 120
Db 61 SDARKSETTITQSGYNGADVQGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGQYGG 120
QY 121 NNAALVNQTASDSSVMVROVGFNNATANQY 151
Db 121 NNPAALVNQTASDSSVMVROVGFNNATANQY 151

```

```

RESULT 5
US-09-543-407-14
; Sequence 14, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 of Leishmania major.
US-09-543-407-14

```

```

Query Match 85.8%; Score 659; DB 19; Length 151;
Best Local Similarity 81.9%; Pred. No. 2.7e-61; Indels 30; Gaps 2;
Matches 136; Conservative 0; Mismatches 0;
QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGSSGDPSTLSIYQYGSANAALALQ 60
Db 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGSSGDPSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVQGGADNSTIELTQNGFRNNAT-----YDQ 105
Db 61 SDARKSETTITQSGYNGADVQGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGQYDQ 120
QY 106 LVTRVVTHEMAHAGGNAALVNQTASDSSVMVROVGFNNATANQY 151
Db 121 LVTRVVTHEMAHA-----SVMVROVGFNNATANQY 151

```

```

RESULT 6
US-09-543-407-22
; Sequence 22, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 22
; LENGTH: 151

```

```

; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-22
Query Match 85.5%; Score 657; DB 19; Length 151;
Best Local Similarity 81.9%; Pred. No. 4.3e-61; Indels 30; Gaps 2;
Matches 136; Conservative 0; Mismatches 0;
QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGSSGDPSTLSIYQYGSANAALALQ 60
Db 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGSSGDPSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVQGGADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHA-- 118
Db 61 SDARKSETTITQSGYNGADVQGGADN-----YDQLVTRVVTHEMAHADQ 105
QY 119 -----GNNNAALVNQTASDSSVMVROVGFNNATANQY 151
Db 106 WNAKNSDITVGQYGGNNAALVNQTASDSSVMVROVGFNNATANQY 151

```

```

RESULT 7
US-09-543-407-28
; Sequence 28, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-28

```

```

Query Match 79.8%; Score 613; DB 19; Length 151;
Best Local Similarity 74.6%; Pred. No. 2e-56; Indels 44; Gaps 2;
Matches 129; Conservative 0; Mismatches 0;
QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGSSGDPSTLSIYQYGSANAALALQ 60
Db 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGSSGDPSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVQGGADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHA-- 118
Db 61 SDARKSETTITQSGYNGAD-----YDQLVTRVVTHEMAHAJR 98
QY 119 -----GNNNAALVNQTASDSSVMVROVGFNNATANQY 151
Db 99 NNATIDOWNAKNSDITVGQYGGNNAALVNQTASDSSVMVROVGFNNATANQY 151

```

```

RESULT 8
US-09-543-407-12
; Sequence 12, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.

```

```
/ APPLICANT: Collinson, S. Karen
/ APPLICANT: Kay, William W.
/ TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
/ PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
/ FILE REFERENCE: 920043.406
/ CURRENT APPLICATION NUMBER: US/09/543,407
/ CURRENT FILING DATE: 2000-04-05
/ NUMBER OF SEQ ID NOS: 59
/ SOFTWARE: FastSEQ for Windows Version 4.0
/ SEQ ID NO 12
/ LENGTH: 151
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
/ OTHER INFORMATION: sequence containing the replacement fragment
/ OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-12

Query Match      78.5%; Score 603; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 2.4e-55;
Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPWGGGNGHNGSGSPDSTLSIYQYGSANNAALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPWGGGNGHNGSGSPDSTLSIYQYGSANNAALQ 60
QY 61 SPARKSETTITQSGYNGADVGGQADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGG 120
DB 61 SPARKSETTITQSGYNGADVGGQADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGG 120
QY 121 NNAALVNOTASDSSVMVQVGFNNATANQY 151
DB 121 NNAALVNDQLVTRVVTHEMAHANNATANQY 151

RESULT 9
US-09-543-407-26
/ Sequence 26, Application US/09543407
/ GENERAL INFORMATION:
/ APPLICANT: White, Aaron P.
/ APPLICANT: Doran, James L.
/ APPLICANT: Collinson, S. Karen
/ APPLICANT: Kay, William W.
/ TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
/ PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
/ FILE REFERENCE: 920043.406
/ CURRENT APPLICATION NUMBER: US/09/543,407
/ CURRENT FILING DATE: 2000-04-05
/ NUMBER OF SEQ ID NOS: 59
/ SOFTWARE: FastSEQ for Windows Version 4.0
/ SEQ ID NO 26
/ LENGTH: 151
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
/ OTHER INFORMATION: sequence containing the replacement fragment
/ OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-26

Query Match      78.1%; Score 600; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 4.9e-55;
Matches 123; Conservative 4; Mismatches 24; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPWGGGNGHNGSGSPDSTLSIYQYGSANNAALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPWGGGNGHNGSGSPDSTLSIYQYGSANNAALQ 60
QY 61 SPARKSETTITQSGYNGADVGGQADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGG 120
DB 61 LVTRVVTHEMAHAGYNGADVGGQADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGG 120
```

```
QY 121 NNAALVNOTASDSSVMVQVGFNNATANQY 151
DB 121 NNAALVNOTASDSSVMVQVGFNNATANQY 151

RESULT 10
US-09-543-407-18
/ Sequence 18, Application US/09543407
/ GENERAL INFORMATION:
/ APPLICANT: White, Aaron P.
/ APPLICANT: Doran, James L.
/ APPLICANT: Collinson, S. Karen
/ APPLICANT: Kay, William W.
/ TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
/ PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
/ FILE REFERENCE: 920043.406
/ CURRENT APPLICATION NUMBER: US/09/543,407
/ CURRENT FILING DATE: 2000-04-05
/ NUMBER OF SEQ ID NOS: 59
/ SOFTWARE: FastSEQ for Windows Version 4.0
/ SEQ ID NO 18
/ LENGTH: 151
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
/ OTHER INFORMATION: sequence containing the replacement fragment
/ OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-18
```

```
Query Match      78.0%; Score 599; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 6.3e-55;
Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPWGGGNGHNGSGSPDSTLSIYQYGSANNAALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPWGGGNGHNGSGSPDSTLSIYQYGSANNAALQ 60
QY 61 SPARKSETTITQSGYNGADVGGQADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGG 120
DB 61 SPARKSETTITQSGYNGADVGGQADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGG 120
QY 121 NNAALVNOTASDSSVMVQVGFNNATANQY 151
DB 121 NNAALVNOTASDSSVMVQVGFNNATANQY 151
```

```
RESULT 11
US-09-543-407-20
/ Sequence 20, Application US/09543407
/ GENERAL INFORMATION:
/ APPLICANT: White, Aaron P.
/ APPLICANT: Doran, James L.
/ APPLICANT: Collinson, S. Karen
/ APPLICANT: Kay, William W.
/ TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
/ PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
/ FILE REFERENCE: 920043.406
/ CURRENT APPLICATION NUMBER: US/09/543,407
/ CURRENT FILING DATE: 2000-04-05
/ NUMBER OF SEQ ID NOS: 59
/ SOFTWARE: FastSEQ for Windows Version 4.0
/ SEQ ID NO 20
/ LENGTH: 151
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
/ OTHER INFORMATION: sequence containing the replacement fragment
/ OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-20
```

Query Match 77.7%; Score 597; DB 19; Length 151;

```
Best Local Similarity 81.5%; Pred. No. 1e-54;
Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALQ 60
QY 61 SDARKSETTITQSGYNGADVGGADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGG 120
Db 61 SDARKYDQLVTRVVTHEMAHAGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
QY 121 NNAALVNOTASDSSVMVROVGFNNATANQY 151
Db 121 NNAALVNOTASDSSVMVROVGFNNATANQY 151

RESULT 12
US-09-543-407-31
; Sequence 31, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543.407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Salmonella enteritidis
US-09-543-407-31

Query Match 77.3%; Score 594; DB 19; Length 131;
Best Local Similarity 89.3%; Pred. No. 1.8e-54;
Matches 117; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

QY 21 GVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALQSDARKSETTITQSGYNGAD 80
Db 1 GVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALQSDARKSETTITQSGYNGAD 60
QY 81 VQGADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGGNNALVNOTASDSSVMVROV 140
Db 61 VQGADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGGNNAALVNOTASDSSVMVROV 120
QY 141 GFNNATANQY 151
Db 121 GFNNATANQY 131

RESULT 13
US-09-543-407-16
; Sequence 16, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543.407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
```

```
FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-16

Query Match 73.7%; Score 566; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 2e-51;
Matches 122; Conservative 5; Mismatches 24; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVYDQLVTRVVTHEMAHAGSPDSTLSIYQYGSANAALQ 60
QY 61 SDARKSETTITQSGYNGADVGGADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGG 120
Db 61 SDARKSETTITQSGYNGADVGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
QY 121 NNAALVNOTASDSSVMVROVGFNNATANQY 151
Db 121 NNAALVNOTASDSSVMVROVGFNNATANQY 151

RESULT 14
US-09-543-407-7
; Sequence 7, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543.407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-543-407-7

Query Match 67.3%; Score 517; DB 19; Length 151;
Best Local Similarity 69.5%; Pred. No. 3.2e-46;
Matches 105; Conservative 17; Mismatches 29; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQYGGGNGHNGSGSPNSLNIYQYGGNSALALQ 60
QY 61 SDARKSETTITQSGYNGADVGGADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGG 120
Db 61 TDARNSDLTITQHGNGGADVGGGDDSSIDLTQRGFGNSATIDQWNGKNSMTVQVQFG 120
QY 121 NNAALVNOTASDSSVMVROVGFNNATANQY 151
Db 121 GNCAAVDQTASNSVNVTVQVGFNNATANQY 151

RESULT 15
US-08-978-878-4
; Sequence 4, Application US/08978878
; GENERAL INFORMATION:
; APPLICANT: NORMARK, Staffan
; APPLICANT: OLSEN, Arne
; TITLE OF INVENTION: FIBRONECTIN BINDING PROTEIN AS WELL AS ITS PREPARATION
; FILE REFERENCE: 012889-081
; CURRENT APPLICATION NUMBER: US/08/978.878
; CURRENT FILING DATE: 1997-11-26
; EARLIER APPLICATION NUMBER: SE 8801723-1
; EARLIER FILING DATE: 1988-05-06
```

```
; EARLIER APPLICATION NUMBER: US 07/347,189
; EARLIER FILING DATE: 1989-05-04
; EARLIER APPLICATION NUMBER: US 07/789,437
; EARLIER FILING DATE: 1991-11-06
; EARLIER APPLICATION NUMBER: US 07/970,846
; EARLIER FILING DATE: 1992-11-03
; EARLIER APPLICATION NUMBER: US 08/187,865
; EARLIER FILING DATE: 1994-01-28
; EARLIER APPLICATION NUMBER: US 08/318,519
; EARLIER FILING DATE: 1994-10-05
; EARLIER APPLICATION NUMBER: US 08/495,959
; EARLIER FILING DATE: 1995-06-28
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-08-978-878-4

Query Match          66.9%; Score 514; DB 13; Length 151;
Best Local Similarity 68.9%; Pred. No. 6.7e-46;
Matches 104; Conservative 18; Mismatches 29; Indels 0; Gaps 0;

QY      1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNNHGGSSGPDSTLSIYQGSANAALALQ 60
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      1 MKLLKVAATAIIVFSGSAVAGVVPQYGGGNNHGGSSGPDSTLSIYQGSANALALQ 60
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY      61 SPARKSETTITQSGYGNAGADVCGGADNSTIELTQNGPRNNATYDCLVTRVVVTHEMAHAGG 120
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      61 TDARNSDLTITQGGGNGADVCGGSDSSIDLITQGFNGSATLDWNGKNSMTVKQFGG 120
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY      121 NNAALVNQTSADSSVMVRQVGFNNATANQY 151
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      121 GNGAAVDQTASNSVNVTVQVGFNNATAHQY 151
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
```

Search completed: August 2, 2004, 15:26:47
Job time : 168.9 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:49:38 ; Search time 17.8 seconds
(without alignments)
888.146 Million cell updates/sec

Title: US-09-543-407-30

Perfect score: 768
Sequence: 1 MKLKVAFAAIVVSGSALA.....DSSVMVQVGFNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 601315 seqs, 104695340 residues

Total number of hits satisfying chosen parameters: 601315

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents AA New:*

1: /cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pep:*
2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep:*
3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep:*
4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep:*
5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:*
6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:*
7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	514	66.9	151	US-09-741-873C-4	Sequence 4, Appli
2	436	56.8	131	US-09-741-873C-2	Sequence 2, Appli
3	101.5	13.2	520	US-10-479-638-21	Sequence 21, Appl
4	93.5	12.2	256	US-10-425-115-301334	Sequence 301334,
5	93	12.1	258	US-10-425-115-300390	Sequence 300390,
6	92	12.0	295	US-10-425-115-312468	Sequence 312468,
7	91.5	11.9	179	US-10-425-115-346132	Sequence 346132,
8	91.5	11.9	299	US-10-170-205E-35751	Sequence 35751, A
9	91.5	11.9	1203	US-10-170-205E-741	Sequence 741, App
10	91.5	11.9	1327	PCT-US04-02338-49	Sequence 49, Appl
11	88.5	11.5	197	US-10-425-115-304391	Sequence 304391,
12	88.5	11.5	388	US-09-248-796A-17306	Sequence 17306, A
13	87.5	11.4	573	US-60-565-632-7807	Sequence 7807, Ap
14	87.5	11.4	573	US-60-579-062-7907	Sequence 7907, Ap
15	86.5	11.3	193	US-10-425-115-254240	Sequence 254240,
16	86	11.2	131	US-60-565-632-11109	Sequence 11109, A
17	86	11.2	131	US-60-579-062-11109	Sequence 11109, A
18	85	11.1	321	PCT-US04-05654-590	Sequence 590, App
19	85	11.1	511	US-10-425-115-320950	Sequence 320950,
20	85	11.1	841	US-60-565-632-7906	Sequence 7906, Ap
21	85	11.1	841	US-60-579-062-7906	Sequence 7906, Ap
22	84.5	11.0	412	US-60-565-632-7905	Sequence 7905, Ap
23	84.5	11.0	412	US-60-579-062-7905	Sequence 7905, Ap
24	84	10.9	556	US-10-425-115-337674	Sequence 337674,
25	84	10.9	586	PCT-US03-24982A-317	Sequence 317, App
26	83.5	10.9	758	US-60-565-632-9194	Sequence 9194, Ap

ALIGNMENTS

RESULT 1

US-09-741-873C-4
; Sequence 4; Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873C-4

Query Match 66.9%; Score 514; DB 5; Length 151;
Best Local Similarity 68.9%; Pred. No. 3.7e-37;
Matches 104; Conservative 18; Mismatches 29; Indels 0; Gaps 0;
Qy 1 MKLKVAFAAIVVSGSALAAGVVPQMGCGGNGGNGGNSGDPSTLSIYQYGSANALALQ 60
Db 1 MKLKVAFAAIVVSGSALAAGVVPQMGCGGNGGNGGNSGDPSTLSIYQYGSANALALQ 60
Qy 61 SPARKSETTITGSGYNGADVGCGADNNTFIETQGFNRNATYDOLVTRVVTHEMAHAGG 120
Db 61 TARNSDLTITGCGGNGADVGCGGDDSDSIDITQGFNSATLDOWNGKNSMTVQFGG 120
Qy 121 NNAALVNOTASDSSVMVQVGFNNATANQY 151
Db 121 GNGAAVDQTASNSVNVTVQVGFNNATANQY 151

83.5 10.9 758 7 US-60-579-062-9194 Sequence 9194, Ap
83 10.8 376 6 US-10-491-733-2 Sequence 2, Appli
29 83 719 5 US-09-248-796A-17559 Sequence 17559, A
30 82.5 10.7 630 7 US-60-581-351-9810 Sequence 9810, Ap
31 82 10.7 279 6 US-10-425-115-343835 Sequence 343835,
32 82 10.7 443 6 US-10-100-683-7608 Sequence 7608, Ap
33 82 10.7 1358 6 US-10-778-804-11 Sequence 11, Appl
34 81.5 10.6 508 6 US-10-425-115-285216 Sequence 285216,
35 81.5 10.6 532 6 US-10-425-115-285214 Sequence 285214,
36 81 10.5 201 6 US-10-425-115-309662 Sequence 309662,
37 81 10.5 376 6 US-10-425-115-343526 Sequence 343526,
38 81 10.5 1127 7 US-60-581-351-1872 Sequence 1872, Ap
39 80.5 10.5 246 6 US-10-854-439-511 Sequence 511, App
40 80.5 10.5 269 6 US-10-425-115-190187 Sequence 190187,
41 80.5 10.5 630 7 US-60-581-351-9806 Sequence 9806, Ap
42 80.5 10.5 631 7 US-60-581-351-9807 Sequence 9807, Ap
43 80.5 10.4 1571 6 US-10-864-138-2 Sequence 2, Appli
44 80 10.4 132 6 US-10-425-115-351875 Sequence 351875,
45 80 10.4 1059 1 PCT-US03-24982A-343 Sequence 343, App

RESULT 2
 US-09-741-873C-2
 ; Sequence 2, Application US/09741873C
 ; GENERAL INFORMATION:
 ; APPLICANT: Normark, Staffan
 ; APPLICANT: Olsen, Arne
 ; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
 ; FILE REFERENCE: 012889-084
 ; CURRENT APPLICATION NUMBER: US/09/741,873C
 ; CURRENT FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: SE 8801723-1
 ; PRIOR FILING DATE: 1998-05-06
 ; PRIOR APPLICATION NUMBER: US 08/978,878
 ; PRIOR FILING DATE: 1997-11-26
 ; PRIOR APPLICATION NUMBER: US 07/347,189
 ; PRIOR FILING DATE: 1989-05-04
 ; PRIOR APPLICATION NUMBER: US 07/789,437
 ; PRIOR FILING DATE: 1991-11-06
 ; PRIOR APPLICATION NUMBER: US 07/970,846
 ; PRIOR FILING DATE: 1992-11-03
 ; PRIOR APPLICATION NUMBER: US 08/187,865
 ; PRIOR FILING DATE: 1994-01-28
 ; PRIOR APPLICATION NUMBER: US 08/318,519
 ; PRIOR FILING DATE: 1994-10-05
 ; NUMBER OF SEQ ID NOS: 11
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 2
 ; LENGTH: 131
 ; TYPE: PRT
 ; ORGANISM: Escherichia coli
 US-09-741-873C-2

Query Match 56.8%; Score 436; DB 5; Length 131;
 Best Local Similarity 65.6%; Pred. No. 1.7e-30;
 Matches 86; Conservative 17; Mismatches 28; Indels 0; Gaps 0;
 QY 21 GVPWGGGNGHNGGNSGPDSTLSIYQGSANAALALQSDARKSETTIQSGYNGAD 80
 DB 1 GVPYQGGGNGHNGGNSGENSELTYQGGNSALALQTDARNSDLTITQHGNGAD 60
 QY 81 VQGGADNSTIETQNGFRNNATYDQLVTRVVTHEMAHAGNNAALVNQTSADSSVMVRQV 140
 DB 61 VQGGDDSSIDLTQRFGNSATLDQWNGKNSMTVKQFGGNGAVIDQTASNSSVAVTQV 120
 QY 141 GFGNNATANQY 151
 DB 121 GFGNNATANQY 131

RESULT 3
 US-10-479-638-21
 ; Sequence 21, Application US/10479638
 ; GENERAL INFORMATION:
 ; APPLICANT: Don A. Roth
 ; APPLICANT: Randolph V. Lewis
 ; APPLICANT: The University of Wyoming
 ; TITLE OF INVENTION: Expression of Spider Silk Proteins in Higher Plants
 ; FILE REFERENCE: WYO.02-0004US
 ; CURRENT APPLICATION NUMBER: US/10/479,638
 ; CURRENT FILING DATE: 2003-12-03
 ; PRIOR APPLICATION NUMBER: PCT/US02/18256
 ; PRIOR FILING DATE: 2002-06-06
 ; PRIOR APPLICATION NUMBER: 60/296,184
 ; PRIOR FILING DATE: 2001-06-06
 ; NUMBER OF SEQ ID NOS: 54
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 21
 ; LENGTH: 520
 ; TYPE: PRT
 ; ORGANISM: Argiope trifasciata
 US-10-479-638-21
 Query Match 13.2%; Score 101.5; DB 6; Length 520;

Best Local Similarity 24.6%; Pred. No. 0.68;
 Matches 34; Conservative 25; Mismatches 66; Indels 13; Gaps 3;
 QY 15 SGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQGSANAALALQSDARKSET-TITQS 73
 DB 347 AGAGAAAAAAGAGAGGCGGYGAGGSS-----ISYGATSSSATSSSTASSRSRGIVTSG 402
 QY 74 GYGNGADVQGGADNSTIETQNGFRNNATYDQLVTRVVTHEMAHAGNNAALVNQTSADS 133
 DB 403 GYGAGAAAGAGAGAAAGAGSYSGSISRLSSAEAVRVSSNICAVASGGASALPG----- 456
 QY 134 SVMVRQVGFNNATANQY 151
 DB 457 --VISNIFSGVSSAGSY 472

RESULT 4
 US-10-425-115-301334
 ; Sequence 301334, Application US/10425115
 ; GENERAL INFORMATION:
 ; APPLICANT: La Rosa, Thomas J.
 ; APPLICANT: Kovalic, David K.
 ; APPLICANT: Zhou, Yihua
 ; APPLICANT: Cao, Yongwei
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
 ; FILE REFERENCE: 38-21(53222)B
 ; CURRENT APPLICATION NUMBER: US/10/425,115
 ; CURRENT FILING DATE: 2003-04-28
 ; NUMBER OF SEQ ID NOS: 369326
 ; SEQ ID NO 301334
 ; LENGTH: 256
 ; TYPE: PRT
 ; ORGANISM: Zea mays
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: MRT4577_37894C.1.pep
 US-10-425-115-301334

Query Match 12.2%; Score 93.5; DB 6; Length 256;
 Best Local Similarity 25.2%; Pred. No. 1.5;
 Matches 38; Conservative 15; Mismatches 51; Indels 47; Gaps 6;
 QY 17 SALAGVVPQWGGGNGHNGGNSGPDSTLSIYQGSANAALALQSDA-----EKSETTIQ 72
 DB 79 SVVAG-----GGGGGGGGGTGGGGGGGGGTGTAASGPGSSGNYANAEGKAG 133
 QY 73 SGYNGAD-----VQGGADNSTIEL--TQNGFRNNATYDQLVTRVVTHEMAHAG 120
 DB 134 GGMGGADGAYGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 182
 QY 121 NNAALVNQTSADSSVMVRQVGFNNATANQY 151
 DB 183 SGAG-----GGHGGGAGAPSY 198

RESULT 5
 US-10-425-115-300390
 ; Sequence 300390, Application US/10425115
 ; GENERAL INFORMATION:
 ; APPLICANT: La Rosa, Thomas J.
 ; APPLICANT: Kovalic, David K.
 ; APPLICANT: Zhou, Yihua
 ; APPLICANT: Cao, Yongwei
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
 ; FILE REFERENCE: 38-21(53222)B
 ; CURRENT APPLICATION NUMBER: US/10/425,115
 ; CURRENT FILING DATE: 2003-04-28
 ; NUMBER OF SEQ ID NOS: 369326
 ; SEQ ID NO 300390
 ; LENGTH: 258
 ; TYPE: PRT
 ; ORGANISM: Zea mays

```
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(258)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_37025C.1.pep
US-10-425-115-300390

Query Match      12.1%; Score 93; DB 6; Length 258;
Best Local Similarity 26.8%; Pred. No. 1.7; 42; Indels 32; Gaps 5;
Matches 33; Conservative 16; Mismatches 16;

QY 17 SALACVPPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQSDARKSETTITQSG-- 74
Db 78 SSIAG-----GGGGGGGGGTNGSGSGGGGSGYGSSTAAAGSPSSGNVADAEKGAG 132
QY 75 -----YNGA--DVGGGANDSTIEL--TQGNFNATYDQVTRVVTTHMAHAGG 120
Db 133 GGMGGGANGAIGSAGGGVGKGEGVGVVALAPSSNGYNGGAAD-----ATGGG 181
QY 121 NNA 123
Db 182 SGA 184

RESULT 6
US-10-425-115-312468
; Sequence 312468, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 312468
; LENGTH: 295
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_48027C.1.pep
US-10-425-115-312468

Query Match      12.0%; Score 92; DB 6; Length 295;
Best Local Similarity 30.2%; Pred. No. 2.4;
Matches 35; Conservative 13; Mismatches 46; Indels 22; Gaps 5;

QY 26 WGGGNGHNGGNSGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYGNADVCGGA 85
Db 138 YGGGYSGGGGYSYG-GYAAAGYGVGSGGNYSNASGGVSGS-----DGYGNGAASGGYA 192
QY 86 DNSTIELTQGNFRNNATYDQVTRVVTTHMAHAGGNN-----AALVNQTASDSS 134
Db 193 NN-----LSSGYNNGRYN-----TIGSDGNTGGYNSYRPNYPGAGNYNTGSSSS 238

RESULT 7
US-10-425-115-346132
; Sequence 346132, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
```

```
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 346132
; LENGTH: 179
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_78839C.1.pep
US-10-425-115-346132

Query Match      11.9%; Score 91.5; DB 6; Length 179;
Best Local Similarity 27.7%; Pred. No. 1.5; 55; Indels 37; Gaps 6;
Matches 39; Conservative 10; Mismatches 10;

QY 19 LAGVPPQWGGG-----NHNGGNSGPDSTLSIYQYGSANAALALQSDARKSETTI 70
Db 64 LDGLLSLGGGLGLDGLGLTGGGGNGKENAQADSGNAQEGSGNA---QEDSGNAQ--- 116
QY 71 TQSGYGNADVGGGANDSTIELTQGNFRNNATYDQVTRVVTTHMAHAGGNNALVN-QT 129
Db 117 EESGNNQNAAGAGAGA-----ENGANNGT-----EAAGAENAAAGNGQA 155
QY 130 ASDSSVMVRQVGVGNNTAANQ 150
Db 156 QIEGEDEQAQGNAGNENAAEE 176

RESULT 8
US-10-170-205E-35751
; Sequence 35751, Application US/10170205E
; GENERAL INFORMATION:
; APPLICANT: ADAMS, Mark
; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN
; FILE REFERENCE: CL001381
; CURRENT APPLICATION NUMBER: US/10/170,205E
; CURRENT FILING DATE: 2002-06-13
; NUMBER OF SEQ ID NOS: 40312
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 35751
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Homo sapiens
; OTHER INFORMATION: Clone ID: MRT4577_48027C.1.pep
US-10-170-205E-35751

Query Match      11.9%; Score 91.5; DB 6; Length 299;
Best Local Similarity 29.4%; Pred. No. 2.6;
Matches 42; Conservative 17; Mismatches 65; Indels 19; Gaps 6;

QY 6 VAAFAAI-VVGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQSDAR 64
Db 99 VAAAPVVPVAVSTSSAAGVAPNPAGSGSNNSPSSSSPTSS-SSSSPSSPGSSLAESPEAA 157
QY 65 KSETTIT-----QSGYGNADVCGGADNSTIELTQGNFRNNATYDQVTRVVTTHMAHAGG 120
Db 158 GVSSTAPLQFGAAGPQTGPVAVSGALRELLAACRNGD-----VSRV--KRLVDAAN 206
QY 121 NNAALVNQTASDSSVMVRQVGF 143
Db 207 VNAAK--DMAGRKSSPLHFAAGFG 227

RESULT 9
US-10-170-205E-741
; Sequence 741, Application US/10170205E
; GENERAL INFORMATION:
; APPLICANT: ADAMS, Mark
; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN
; FILE REFERENCE: CL001381
; CURRENT APPLICATION NUMBER: US/10/170,205E
; CURRENT FILING DATE: 2002-06-13
; NUMBER OF SEQ ID NOS: 40312
; SOFTWARE: Patent in version 3.2
```


Db 240 GAGFGDNNTSSY 252

RESULT 13

US-60-565-632-7907

; Sequence 7907, Application US/60565632

; GENERAL INFORMATION:

; APPLICANT: Monsanto Technology, LLC

; APPLICANT: Baum, James A

; APPLICANT: Kovalic, David K.

; APPLICANT: Larosa, Thomas J.

; APPLICANT: Lu, Maolong

; APPLICANT: Munyikwa, Tichifa R. I.

; APPLICANT: Roberts, James K.

; APPLICANT: Wu, Wei

; APPLICANT: Zhang, Bei

; TITLE OF INVENTION: Methods for Genetic Control of Insect Infestations in Plants and

; TITLE OF INVENTION: Compositions Thereof

; FILE REFERENCE: 38-21(53403)B

; CURRENT APPLICATION NUMBER: US/60565,632

; CURRENT FILING DATE: 2004-04-27

; NUMBER OF SEQ ID NOS: 15449

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 7907

; LENGTH: 573

; TYPE: PRT

; ORGANISM: Diabrotica virgifera

US-60-565-632-7907

Query Match 11.4%; Score 87.5; DB 7; Length 573;

Best Local Similarity 26.0%; Pred. NO.12;

Matches 34; Conservative 16; Mismatches 54; Indels 27; Gaps 5;

QY 30 GNHHGGG-NSSGGPDSLISYQVGSAAALA-----LQSDARKSETTITQSGVGNAD 80

Db 241 GNENGTGAENNVANADACTDVAQ-GSTNEAENNVANADVQNDAAQENNGAAENSGNAD 299

QY 81 VQGGADN-STIELTON-----GFRNATYQQLVTRVVVTHEMAHAGNNA 123

Db 300 AAQCTDNGAAENGTGNADPAQGNNGAAENSGNENGTAENNVANADVQNDAAQVNDNGA 359

QY 124 ALVNQATSDSS 134

Db 360 AENNGNADAA 370

RESULT 14

US-60-579-062-7907

; Sequence 7907, Application US/60579062

; GENERAL INFORMATION:

; APPLICANT: Baum, James A

; APPLICANT: Kovalic, David K

; APPLICANT: Larosa, Thomas J

; APPLICANT: Lu, Maolong

; APPLICANT: Munyikwa, Tichifa R. I.

; APPLICANT: Roberts, James K

; APPLICANT: Wu, Wei

; APPLICANT: Zhang, Bei

; TITLE OF INVENTION: Methods for Genetic Control of Insect Infestations in Plants and

; TITLE OF INVENTION: Compositions thereof

; FILE REFERENCE: 38-21 (53403) C

; CURRENT APPLICATION NUMBER: US/60579,062

; CURRENT FILING DATE: 2004-06-11

; NUMBER OF SEQ ID NOS: 41445

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 7907

; LENGTH: 573

; TYPE: PRT

; ORGANISM: Diabrotica virgifera

US-60-579-062-7907

Query Match	11.4%	Score 87.5;	DB 7;	Length 573;
Best Local Similarity	26.0%	Pred. No. 12;		

```

Matches 34; Conservative 16; Mismatches 54; Indels 27; Gaps 5;
QY 30 GNHNGG--NSGGPDSTLSIYQGSANAALA-----LQSDARKSETTITQSGYNGAD 80
Db 241 GNENGTGAENANADAQTDVAQ-GSTNEAENANANADVQNDAAQANENGAAENSGNAD 299
QY 81 VQGGADN-STIELTON-----GFENNATYDQLVTRVVTHEMAHAGGNA 123
Db 300 AAQGTNDGAAAEENTGNADPAQGNNDNGAAAEENSGNENGTAENANANADVQNDAAQVNDNGA 359
QY 124 ALVNQTASDSS 134
Db 360 AAENNGNADAA 370

RESULT 15
US-10-425-115-254240
; Sequence 254240, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222) B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 254240
; LENGTH: 193
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_163446C.1.pep
US-10-425-115-254240

Query Match 11.3%; Score 86.5; DB 6; Length 193;
Best Local Similarity 26.9%; Pred. No. 4.3;
Matches 35; Conservative 15; Mismatches 53; Indels 27; Gaps 5;
QY 27 GGGGNHNGG--GNSSGPDSTLSIYQGSANAALALQSDARKSETTITQSGYNGADYQ 83
Db 76 GGGASQNGSGSGYSGSGSGSGSTYSQGGYISGYGESSNA-----GGTGGGGGGGQ 126
QY 84 --GADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGGNAALVNQT-----ASDSS 134
Db 127 AGGYWYNNAQCGSGGTGGSSYANRYWDGSGEGGANANGGGMGNGQNGGGGGSGDGS 186
QY 135 VMVRQVGFGN 144
Db 187 -----GYGN 190

Search completed: August 2, 2004, 15:29:56
Job time : 18.8 secs

```

Search completed: August 2, 2004, 15:29:56
Job time : 18.8 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:39:53 ; Search time 9.4 Seconds

(without alignments)
1545.204 Million cell updates/sec

Title: US-09-543-407-30

Perfect score: 768

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVGFNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78:*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	681	88.7	151	2 JC6039	fimbrin protein ag
2	681	88.7	151	2 A10635	major curlin chain
3	517	67.3	151	2 S70788	curlin protein csg
4	502.5	65.4	152	2 D30806	curlin major subun
5	502.5	65.4	152	2 H85665	hypothetical prote
6	104.5	13.6	151	2 S70787	curlin nucleator p
7	104.5	13.6	151	2 C30806	minor curlin subun
8	104.5	13.6	151	2 G85665	curlin minor chain
9	101	13.2	1748	2 S42136	cnjB protein - Tet
10	99.5	13.0	151	2 JC6040	fimbrin protein ag
11	99.5	13.0	151	2 AH0635	nucleation compone
12	99.5	13.0	2174	2 E35965	hypothetical glyci
13	98	12.8	145	2 AD3143	conserved hypohet
14	98	12.8	145	2 H98144	hypothetical prote
15	97.5	12.7	645	2 F70825	probable PPE prote
16	96.5	12.6	313	2 T04776	hypothetical prote
17	94	12.2	354	2 B70663	probable PPE prote
18	94	12.2	656	2 T96831	hypothetical prote
19	93.5	12.2	256	2 T03371	glycine-rich prote
20	93.5	12.2	2204	2 A70534	probable PPE prote
21	92.5	12.0	1028	2 A56038	DNA-binding protei
22	92.5	12.0	1213	2 S16356	ovo protein - fru1
23	92	12.0	575	2 S35327	protein kinase sgg
24	91.5	11.9	409	2 T20847	hypothetical prote
25	91.5	11.9	573	2 C86266	F3f19.21 protein -
26	91.5	11.9	1910	2 AF0394	probable adhesin h
27	91	11.8	652	2 E97857	cell surface antig
28	90.5	11.8	447	2 G84657	probable disease r
29	89.5	11.7	963	2 B70524	probable PPE prote

30 89.5 11.7 1238 2 AH0038 probable exported
31 89 11.6 343 2 T05221 hypothetical prote
32 89 11.6 959 2 B44402 nuclear pore compl
33 88.5 11.5 967 2 S66852 hypothetical prote
34 88.5 11.5 1217 2 S52714 sericin1B - silkw
35 88 11.5 407 2 T21956 hypothetical prote
36 88 11.5 586 2 T26667 hypothetical prote
37 88 11.5 1778 2 T50074 probable nucleopor
38 87.5 11.4 1053 2 B70987 probable PPE prote
39 87 11.3 3624 2 AD0835 large repetitive p
40 86.5 11.3 262 2 S00275 tail fiber protein
41 86.5 11.3 590 1 A45621 leishmanolysin (EC
42 86.5 11.3 745 2 E64559 leishmanolysin (EC
43 86.5 11.3 745 2 E64559 outer membrane pro
44 85 11.2 582 2 F70675 probable PPE prote
45 85 11.2 639 2 C42049 leishmanolysin (EC

ALIGNMENTS

RESULT 1

JC6039
fimbrin protein agfa precursor - Salmonella enteritidis

C:Species: Salmonella enteritidis

C>Date: 31-Dec-1996 #sequence revision 31-Dec-1996 #text_change 08-Oct-1999

C:Accession: JC6039; PC6015; A44898

R:Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Baner, P.A.; Kay, W.W.

J. Bacteriol. 178, 662-667, 1996

A:Title: Salmonella enteritidis agfA operon encoding thin, aggregative fimbriae.

A:Reference number: JC6039; MUID:96146512; PMID:8550497

A:Accession: JC6039

A:Molecule type: DNA

A:Residues: 1-151 <COL>

A:Cross-references: GB:U43280; NID:gil184712; PIDN:AAC43599.1; PID:gil184714

A:Accession: PC6015

A:Molecule type: Protein

A:Residues: 21-52 <CO2>

A:Experimental source: strain 27655-3b

A:Note: the authors translated the codon ACG for residue 44 as Ile

R:Collinson, S.K.; Emody, L.; Muller, K.H.; Trust, T.J.; Kay, W.W.

J. Bacteriol. 173, 4773-4781, 1991

A:Title: Purification and characterization of thin, aggregative fimbriae from Salmonella

A:Reference number: A44898; MUID:91310586; PMID:1677357

A:Contents: 27655

A:Accession: A44898

A>Status: preliminary

A:Molecule type: protein

A:Residues: 21-33 <CO3>

A:Note: sequence extracted from NCBI backbone (NCBIP:45936)

C:Genetics:

A:Gene: agfA

C:Function:

A:Description: major component of thin aggregative fimbriae

A:Note: fimbriae bind to fibronectin, plasminogen, tissue plasminogen activator

C:Keywords: fimbrin

F:1-20/Domain: signal sequence #status predicted <SIG>

F:21-151/Product: fimbrin protein agfA #status experimental <MAT>

Query Match 88.7%; Score 681; DB 2; Length 151;

Best Local Similarity 90.7%; Pred. No. 3.3e-49;

Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGNGSGSPDSTLSIYQGSANAALALQ 60

Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGNGSGSPDSTLSIYQGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGADVQGGADNSTIETQTQGFNNATYDOLVTRVVTHEMAHAGG 120

Db 61 SDARKSETTITQSGYNGADVQGGADNSTIETQTQGFNNATIDQWNAKNSDIIVQYGG 120

QY 121 NNAALVNQATSSVVRQVGFNNATANQY 151

Db 121 NNAALVNQASDSSVMVRQVGFNNATANQY 151

RESULT 2

A:0635

major curlin chain precursor [imported] - Salmonella enterica subsp. enterica serovar Typhimurium

C:Species: Salmonella enterica subsp. enterica serovar Typhi

A:Note: This species has also been called Salmonella typhi

C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002

C:Accession: A10635

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Churcher, C.; Holt, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, M.; Mouton, S.; O'Gaora, P.

A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhimurium

A:Reference number: AB0502; UID:21534947; PMID:11677608

A:Accession: A10635

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-151 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD08268.1; PID:G16502315; GSPDB:GN00176

C:Genetics:

A:Gene: STY1181

Query Match 88.7%; Score 681; DB 2; Length 151;

Best Local Similarity 90.7%; Pred. No. 3.3e-49;

Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60

Db 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60

QY 61 SPARKSETTITQSGYNGADVCGGADNSTIELTQNGFRNATYDQLVTRVTHEMAHAGG 120

Db 61 SPARKSETTITQSGYNGADVCGGADNSTIELTQNGFRNATYDQLVTRVTHEMAHAGG 120

QY 121 NNAALVNQASDSSVMVRQVGFNNATANQY 151

Db 121 NNAALVNQASDSSVMVRQVGFNNATANQY 151

RESULT 3

S70788

curlin protein csgA precursor - Escherichia coli (strain K-12)

N:Alternate names: csgA protein; major curlin protein

C:Species: Escherichia coli

C:Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 01-Mar-2002

C:Accession: S70788; G64846; S34559

R:Hammar, M.; Arngqvist, A.; Bian, Z.; Olsen, A.; Normark, S.

Mol. Microbiol. 18, 661-670, 1995

A:Title: Expression of two csg operons is required for production of fibronectin- and curli

A:Reference number: S70783; UID:96414468; PMID:8817489

A:Accession: S70788

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-151 <HAM>

A:Cross-references: EMBL:X90754; NID:G1147558; PIDN:CAA62282.1; PID:G1147564

A:Experimental source: strain K12, substrain W3110

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, D.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; UID:97426617; PMID:9278503

A:Accession: G64846

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-151 <BLAT>

A:Cross-references: GB:AE000205; GB:U00096; NID:G1787265; PIDN:AACT4126.1; PID:G1787279;

A:Experimental source: strain K-12, substrain MG1655

R:Olsen, A.; Arngqvist, A.; Hammar, M.; Sukupolvi, S.; Normark, S.

Mol. Microbiol. 7, 523-536, 1993

A:Title: The RpoS sigma factor relieves H-NS-mediated transcriptional repression of csgA

A:Reference number: S31202; UID:93211294; PMID:8459772

A:Accession: S31202

A:Molecule type: DNA

A:Residues: 1-6, 'V', 8-151 <OLS1>

A:Cross-references: EMBL:L04979

A:Accession: S34560

A:Molecule type: protein

A:Residues: 21-42; 44-50 <OLS2>

R:Olsen, A.N.; Arngqvist, A.M.

submitted to the EMBL Data Library, October 1992

A:Reference number: S34559

A:Accession: S34559

A:Molecule type: DNA

A:Residues: 1-133, 'RQRSGMLW' <OLS3>

A:Cross-references: EMBL:L04979; NID:G290424; PIDN:AAA23616.1; PID:G290425

A:Experimental source: strain K-12, substrain W3110

C:Genetics:

A:Gene: csgA

A:Map position: 23.15

C:Function:

A:Description: major component of wild-type curli; interaction between CsgA and CsgB triggers curli assembly

A:Note: curli are thin, coiled fibers expressed on the surface of Escherichia coli that and H-kininogen; in the absence of CsgA, CsgB can self-assemble into polymers

F:A-20/Domain: signal sequence #status predicted <SIG>

F:21-151/Product: curlin #status experimental <MAT>

Query Match 67.3%; Score 517; DB 2; Length 151;

Best Local Similarity 69.5%; Pred. No. 1e-35;

Matches 105; Conservative 17; Mismatches 29; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60

Db 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60

QY 61 SPARKSETTITQSGYNGADVCGGADNSTIELTQNGFRNATYDQLVTRVTHEMAHAGG 120

Db 61 SPARKSETTITQSGYNGADVCGGADNSTIELTQNGFRNATYDQLVTRVTHEMAHAGG 120

QY 121 NNAALVNQASDSSVMVRQVGFNNATANQY 151

Db 121 NNAALVNQASDSSVMVRQVGFNNATANQY 151

RESULT 4

D90806

curlin major subunit CsgA [imported] - Escherichia coli (strain O157:H7, substrain RIMD 00806)

C:Species: Escherichia coli

C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001

C:Accession: D90806

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.; Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic islands

A:Reference number: A99629; UID:21156231; PMID:11258796

A:Accession: D90806

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-152 <HAY>

A:Cross-references: GB:BA000007; PIDN:BA834843.1; PID:G13360880; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RIMD 050952

C:Genetics:

A:Gene: Ecs1420

Query Match 65.4%; Score 502.5; DB 2; Length 152;

Best Local Similarity 68.4%; Pred. No. 1.6e-34;

Matches 104; Conservative 18; Mismatches 29; Indels 1; Gaps 1;

QY 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALAL 59

Db 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALAL 60

QY 60 QSDARKSETTITQSGYNGADVCGGADNSTIELTQNGFRNATYDQLVTRVTHEMAHAG 119

```
Db 61 QADARNSDLTITQGGNGADVGQSGDDSSIDLTRQFGNSATLDQWNGKDSHMTVKQF 120
Qy 120 GNNAAALVNQTASDSVMVROVQFGNNATANQY 151
Db 121 GCGAAAVDQTASNTSVNTVQVGFNNATANQY 152

RESULT 5
H5665
hypothetical protein csgA [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
C:Species: Escherichia coli
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: H85665
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouisis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: H85665
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-152 <STO>
A:Cross-references: GB:AE005174; NID:gl2514574; PIDN:AAG55788.1; GSPDB:GN00145; UWGP:Z16
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: csgA

Query Match 65.4%; Score 502.5; DB 2; Length 152;
Best Local Similarity 68.4%; Pred. No. 1.6e-34;
Matches 104; Conservative 18; Mismatches 29; Indels 1; Gaps 1;

Qy 1 MLLKVAFAAIVVSGSALAGVPOW-GGGNHNGGNSGPDSTLSIYQVGSANAALAL 59
Db 1 MLLKVAFAAIVVSGSALAGVPOYGGGNGHGGGNSGPNSEINIIYQVGGNSALAL 60

Qy 60 QSDARKSETTITQSGYNGADVGQGDNSSTIELTQNGFRNATYDQLVTRVVTHEMAHAG 119
Db 61 QADARNSDLTITQGGNGADVGQSGDDSSIDLTRQFGNSATLDQWNGKDSHMTVKQF 120

Qy 120 GNNAAALVNQTASDSVMVROVQFGNNATANQY 151
Db 121 GCGAAAVDQTASNTSVNTVQVGFNNATANQY 152

RESULT 6
S70787
curlin nucleator protein csgB precursor - Escherichia coli (strain K-12)
N:Alternate names: csgB protein; curlin nucleation component; minor curlin protein
C:Species: Escherichia coli
C>Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 01-Mar-2002
C:Accession: S70787; F64846
R:Hammar, M.; Arngvist, A.; Bian, Z.; Olsen, A.; Normark, S.
Mol. Microbiol. 18, 661-670, 1995
A:Title: Expression of two csg operons is required for production of fibronectin- and Cd
A:Reference number: S70783; MUID:96414468; PMID:8817489
A:Accession: S70787
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-151 <HAM>
A:Cross-references: EMBL:X90754; NID:gl147558; PIDN:CAA62281.1; PID:gl147563
A:Experimental source: strain K12, substrain W3110
A>Note: The nucleotide sequence was submitted to the EMBL Data Library, August 1995
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: F64846
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-151 <BLAT>
A:Cross-references: GB:AE000205; GB:U00096; NID:gl787265; PIDN:AAC74125.1; PID:gl787278;
```

```
A:Experimental source: strain K-12, substrain MG1655
C:Genetics:
A:Gene: csgB
A:Map position: 23.15
C:Function:
A:Description: minor component of wild-type curli; interaction between CsgA and CsgB tr
and H-kininogen; in the absence of CsgA, CsgB can self-assemble into polymers
F;1-21/Domains: signal sequence #status predicted <SIG>
F;22-151/Product: minor curlin chain #status predicted <MAT>

Query Match 13.6%; Score 104.5; DB 2; Length 151;
Best Local Similarity 29.1%; Pred. No. 0.087;
Matches 34; Conservative 15; Mismatches 57; Indels 11; Gaps 3;

Qy 38 SSGPDSSTLSIYQVGSANAALALQSDARKSETTITQSGYNGADVGQGDNSSTIELTQNGF 97
Db 21 AAGDYDLANSEYNF----AVNELSKSFNQAAIIGQAGTNNSAQLRQGGSKLLAVVAQEGS 76

Qy 96 RNNATYDQLVTRVVTHEMAH---AGGNNAAALVNQTASDSVMVROVQFGNNATANQY 151
Db 77 SNRAKIDQ----TGDYNLAYIDQAGSANDASISQAGYNTAMIIQKSGSKANITQY 129

RESULT 7
C90806
minor curlin subunit precursor CsgB [imported] - Escherichia coli (strain O157:H7, subs
C:Species: Escherichia coli
C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: C90806
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G
Gawazara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen
A:Reference number: A95629; MUID:21156231; PMID:11258796
A:Accession: C90806
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-151 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA034842.1; PID:gl33360875; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetics:
A:Gene: ECS1419

Query Match 13.6%; Score 104.5; DB 2; Length 151;
Best Local Similarity 29.1%; Pred. No. 0.087;
Matches 34; Conservative 15; Mismatches 57; Indels 11; Gaps 3;

Qy 38 SSGPDSSTLSIYQVGSANAALALQSDARKSETTITQSGYNGADVGQGDNSSTIELTQNGF 97
Db 21 AAGDYDLANSEYNF----AVNELSKSFNQAAIIGQAGTNNSAQLRQGGSKLLAVVAQEGS 76

Qy 98 RNNATYDQLVTRVVTHEMAH---AGGNNAAALVNQTASDSVMVROVQFGNNATANQY 151
Db 77 SNRAKIDQ----TGDYNLAYIDQAGSANDASISQAGYNTAMIIQKSGSKANITQY 129

RESULT 8
G85665
curlin minor chain precursor, CsgA homolog [imported] - Escherichia coli (strain O157:H
C:Species: Escherichia coli
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: G85665
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhe
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouisis, K.; Apodaca
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: G85665
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-151 <STO>
A:Cross-references: GB:AE005174; NID:gl2514573; PIDN:AAG55787.1; GSPDB:GN00145; UWGP:Z1
```

α₂ NGF KNNK11DQF

QY

Db 74 EGENNRKVDGAGNYNAY-IEQ7GNANDASISQAYGNSAAIIQKSGNKANITQY 129

RESULT 12

E95965

hypothetical glycine-rich protein [imported] - Sinorhizobium meliloti (strain 1021) mag

C:Species: Sinorhizobium meliloti

C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001

C:Accession: E95965

E:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan

P:Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001

A:Title: The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing endo

A:Reference number: A95842; MUID: 21396508; PMID:11481431

A:Accession: E95965

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-2174 <XUR>

A:Cross-references: GB:AL591985; PIDN:CA049389.1; PID:g15140875; GSFDB:GN00167

A:Experimental source: strain 1021, megaplasmid pSymB

R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler, P.;

pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;

Hyman, R.W.; Jones, T.

Science 293, 668-672, 2001

A:Authors: Kahn, D.; Kahn, M.L.; Kalkman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,

hebaült, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yen, K.

A:Title: the composite genome of the legume symbiont Sinorhizobium meliloti.

A:Reference number: A96039; MUID: 21368234; PMID:11474104

C:Genetics: annotation

C:Genetics:

A:Gene: SMB21548

A:Genome: plasmid

Query Match	13.0%;	Score	99.5;	DB	2;	Length	2174;
Best Local Similarity	23.0%;	Pred. No.	4.3;				
Matches	32;	Conservative	14;	Mismatches	66;	Indels	27;
Gaps	2						

QY	11	ATVSGSALAGVVPQGGGNNHGGHSSGDPDSTLSIYQGSANAALAQSDARKSETTI	70
DB	892	AVFTTGDSSHLLAIGSGGGNGGSASS-----MSSTAGTVDITIS	932
QY	71	TOSGCGNADVGQGDADNSTIETLTQNGFRNNATYDQLVTRVVTTHMAHAGNNNAALVNQTA	130
DB	933	ADIGVSVGSGSAGNGGAVTAAUKDSASVTTADDYANAIVQSIGGGGNGGV-----	986
QY	131	SDSSVMVRQVFGNNATAN	149
DB	987	--GSVNSKEIGSGFNLTAN	1003

RESULT 13
AD3143 conserved hypothetical protein Atu4768 [imported] - Agrobacterium tumefaciens (strain C58)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C:Accession: AD3143
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.; Karp, P.; Romero, P.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClellan, S.; Karp, P.; Romero, P.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClellan, S. Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, S.; E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AD3143
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-145 <KUR>
A:Cross-references: GB:AE008689; PIDN:AA14562.1; FID:g17743277; GSPDB:GN00187
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atu4768
A:Map position: linear chromosome

Query Match	12.8%;	Score 98;	DB 2;	Length 145;
Best Local Similarity	27.1%;	Pred. No. 0.29;		
Matches 38;	Conservative 25;	Mismatches 57;	Indels 20;	Gaps 5
Qy	3	LLKVAFAAIVVSGSALAGVP-----QWG-----GGNHNGGN-----SSGPDST	44	
Db	1	MIKSTFASALVALVGLSAAPAMANDVRIEQVGSNSAGAOEGYNRIITVQNGYNR	60	
Qy	45	LSIYQYGSANAALQSDARKSETTITQSYGNGADYVGGADNSTIELTQNGFRNNATYD	104	
Db	61	IVGHQYGRNLS-AVQGEHDNYGSTIQNGNRNVAGIQSGSNHTTILTDQNGNTAAGV	119	
Qy	105	QLVTRVVTTHMAHAGNNAA	124	
Db	120	Q-VGRGCSANVSGGNDNVA	138	

RESULT 14
H98144
hypothetical protein AGR_L 228 [imported] - Agrobacterium tumefaciens (strain
C:Species: Agrobacterium tumefaciens
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
C:Accession: H98144
R:Goodner, B.; Winkler, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.
A.: Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Ma
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobact
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Accession: H98144
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-145 <XUR>
A:Cross-references: GB:AE007870; PTDN:AAK86682.1; PID:g15158413; GSPDB:GN00170
C:Genetics:
A:Gene: AGR_L 228
A:Map position: linear chromosome

```

Query Match      12.8%; Score 98; DB 2; Length 145;
Best Local Similarity 27.1%; Pred. No. 0.29;
Matches 38; Conservative 25; Mismatches 57; Indels 20; Gaps 5

QY      3 LLKVAEAAIVVSGSALAGVVP-----QWG-----GGNHNGCN-----SSGPDS 44
Db      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      45 LGIYGYGANAALALQDARKSETTITQSGYGNAGDVGGADNSTELTQNGFRNNATVD 104
Db      61 IVGHQYGRHNL- AVQGECHDNYGSTTQNGNRNVAGIGQFGSNHITLTQDNGNIAAGV 119

QY      105 QLVTRVVTWTHMAHAGGNAA 124
Db      120 Q-VGRGCSANYSGGNDNVA 138

```

RESULT 15
F70825
probable PPE:protein - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 15-Sep-2003
C:Accession: F70825
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.;
C:Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Genies, S.; Hamlin, N.; Ho-
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, S.; Skelton, S.; Squares, S.;
Nature 393, 537-544, 1998
A:Authors: Agares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete
A:Reference number: A70500; MUID:9295987; PMID:9634230
A:Accession: F70825
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-645 <COL>
A:Cross-references: GB:AL021959; GB:AL123456; NID:93261536; PIDN:CAAL7522.1; PT

A/Resolutions: 1-843 <COL>
A/Cross-references: GB:AL021958; GB:AL123456; NID:g3261536; PIDN:CAA17522.1; PID:e125329

Tue Aug 3 10:54:46 2004

A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: PPE

Query Match 12.7%; Score 97.5; DB 2; Length 645;
Best Local Similarity 25.9%; Pred. No. 1.6;
Matches 43; Conservative 15; Mismatches 59; Indels 49; Gaps 9;

QY 15 SGSALAGVWPO-----WGGGHNHNGGNSGGPDSTLSIYQYGSANAALALQSDARKSEIT 69
DB 237 SGNVVGIGPSSFNVVGSGNIGNANVGGNSG-DNNFGNFGNANIGIGNAGPNMSSPAV 295
QY 70 ITQS-----GYNGADVGGADNSTIELTON-----GFRNATYDQLVTRVVT 112
DB 296 PTPGNGNVGIGNGGNGNFGGN-----TGNANIGLGNVGDGVGFGNSGSYN----- 342
QY 113 HEMAHAGGNAAL-----VNQTA-----SDSSVMVRQVGFGNNTAN 149
DB 343 FGFGNTGNNIGIGLTGSNQIGFGLNSGSG-----NIGFGNSGTGN 384

Search completed: August 2, 2004, 14:56:26
Job time : 10.4 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:36:12 ; Search time 5.3 Seconds
(without alignments)
1483.508 Million cell updates/sec

Title: US-09-543-407-30

Perfect score: 768

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVGFNNATANYQ 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	681	88.7	151	1	CSGA_SALTY
2	517	67.3	151	1	CSGA_ECOLI
3	502.5	65.4	152	1	CSGA_ECO57
4	104.5	13.6	151	1	CSGB_ECOLI
5	99.5	13.0	151	1	CSGB_SALTY
6	99.5	13.0	151	1	CSGB_SALTY
7	92.5	12.0	1028	1	OYO_PROM
8	91.5	11.9	1327	1	TNK1_HUMAN
9	90	11.7	1656	1	OMP_RICUA
10	89	11.6	959	1	N100_YEAST
11	88	11.5	1778	1	N189_SCHPO
12	86.5	11.3	262	1	V338_BPT2
13	86.5	11.3	590	1	GP63_LEIDO
14	86	11.2	646	1	GP63_LEIME
15	85.5	11.1	1567	1	ICEN_XANCT
16	85	11.1	599	1	GP63_LEICH
17	85	11.1	602	1	GP63_LEIMA
18	85	11.1	850	1	PSN_PIEBA
19	85	11.1	1655	1	OMP_RICUN
20	84.5	11.0	678	1	YF48_MYCTU
21	84.5	11.0	760	1	Y31L_ECOLI
22	84.5	11.0	1196	1	ICBV_PSEX
23	84.5	11.0	1469	1	ICLV_PSEX
24	84	10.9	363	1	OMP_KLEPN
25	84	10.9	576	1	ICAF_DROME
26	83.5	10.9	1148	1	ICER_DROME
27	83	10.8	401	1	YK03_CABEL
28	82.5	10.7	592	1	CEA_CITFR
29	81.5	10.6	172	1	CH18_DROME
30	81.5	10.6	347	1	MSA2_PLAF2
31	81.5	10.6	392	1	HWE1_HUMAN
32	81	10.5	165	1	G3P1_ORYSA
33	81	10.5	1845	1	Z236_HUMAN

34	80.5	10.5	534	1	GCR2_YEAST
35	80.5	10.5	874	1	AL56_AZOVI
36	80.5	10.5	880	1	GUN4_THEFU
37	80.5	10.5	1210	1	ICEN_PSEFL
38	80.5	10.5	1258	1	ICEN_ERWHE
39	80	10.4	380	1	PLXA_COLGL
40	80	10.4	491	1	YK98_MYCTU
41	80	10.4	850	1	PRSN_PIEBR
42	80	10.4	1571	1	C3G_DROME
43	80	10.4	5560	1	SPEN_DROME
44	79.5	10.4	163	1	HCF_NATPH
45	79.5	10.4	663	1	DUS8_MOUSE

ALIGNMENTS

RESULT 1
CSGA_SALTY
ID CSGA_SALTY STANDARD; PRT; 151 AA.
AC P55225;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Major curlin subunit precursor (Fimbrin SEF1).
GN CSGA OR AGFA OR STM1144 OR STY1181 OR T1776.
OS Salmonella typhimurium,
OS Salmonella typhi, and
OS Salmonella enteritidis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602, 601, 592;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=SR-11;
RX MEDLINE=98117058; PubMed=9457880;
RA Romling U., Bian Z., Hammar M., Sierralta W.D., Normark S.;
RT "Curli fibers are highly conserved between Salmonella typhimurium and
RT Escherichia coli with respect to operon structure and regulation.";
RL J. Bacteriol. 180:722-731(1998).
[2]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Milvanev E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
[3]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Garra P., Parry C.,
RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrett B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
[4]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=TY2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.P., Rose D.J.,
RA Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;


```
CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L04979; AAA23616.1; -
DR EMBL; X90754; CAA62282.1; -
DR EMBL; AE000205; AAC74126.1; -
DR EMBL; D90741; BAA35832.1; -
DR EMBL; D90742; BAA35840.1; -
DR PIR; S70788; S70788.
DR Ecocore; Egl1489; csGA.
KW Fibria; Signal; Complete proteome.
FT SIGNAL 1 20
FT CHAIN 21 151 MAJOR CURLIN SUBUNIT.
FT CONFLICT 7
FT SEQUENCE 151 AA; 15049 MW; C003470D208D395F CRC64;
Query Match 67.3%; Score 517; DB 1; Length 151;
Best Local Similarity 69.5%; Pred. No. 9.7e-37;
Matches 105; Conservative 17; Mismatches 29; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGNGHNGSGSPDSTLSIYQYGSNAALALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQYGGGNGHNGSGNSPNSLNIYQYGGNSALALQ 60
QY 61 SDARKSETTITQSGYNGADYVGOGADNSTIELTQNGFRNNATYDQLVTRVVTTHMAHAG 120
DB 61 TDARNSDLTITQHGCGNGADYVGQSDSDSIDLTQGFNGSATLDQWNGKNSMTVKQFG 120
QY 121 NNAALVNOTASDSSVMYRVQVFGNNATANY 151
DB 121 GNGAAVDQTASNSTVNTVQVFGNNATAHQY 151
RESULT 3
CSGA_ECO57 STANDARD; PRT; 152 AA.
AC Q93U24;
RC STRAIN=O157:H7 / ATCC 43895;
RX MEDLINE=21218556; PubMed=11319125;
RA Ulrich G.A., Keen J.E., Elder R.O.;
RT "Mutations in the csGP promoter associated with variations in curli
RT expression in certain strains of Escherichia coli O157:H7."
RL Appl. Environ. Microbiol. 67:2367-2370(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
RL Nature 409:529-533(2001).
```

```
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kihara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12."
RL DNA Res. 8:11-22(2001).
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN.
CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; A275733; AAK53212.1; -
DR EMBL; AE005315; AAG55788.1; -
DR EMBL; AP002554; BAB34843.1; -
DR PIR; D90806; D90806.
DR PIR; H85665; H85665.
KW Fibria; Signal; Complete proteome.
FT SIGNAL 1 20
FT CHAIN 21 152 MAJOR CURLIN SUBUNIT.
FT SEQUENCE 152 AA; 15099 MW; EE2D294DDE91243 CRC64;
Query Match 65.4%; Score 502.5; DB 1; Length 152;
Best Local Similarity 68.4%; Pred. No. 1.6e-35;
Matches 104; Conservative 18; Mismatches 29; Indels 1; Gaps 1;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQW-GGGNGHNGSGSGSPDSTLSIYQYGSNAALAL 59
DB 1 MKLLKVAFAAIVVSGSALAGVVPQYGGGNGHNGSGNSPNSLNIYQYGGNSALAL 60
QY 60 QSDARKSETTITQSGYNGADYVGOGADNSTIELTQNGFRNNATYDQLVTRVVTTHMAHAG 119
DB 61 QADARNSDLTITQHGCGNGADYVGQSDSDSIDLTQGFNGSATLDQWNGKNSMTVKQFG 120
QY 120 GNAALVNOTASDSSVMYRVQVFGNNATANY 151
DB 121 GNGAAVDQTASNSTVNTVQVFGNNATAHQY 152
RESULT 4
CSGB_ECOLI STANDARD; PRT; 151 AA.
ID CSGB_ECOLI
AC F39828;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Minor curlin subunit precursor.
GN CSGB OR B1041 OR Z1675 OR ECS1419.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=X12 / MC4100;
RX MEDLINE=96414468; PubMed=8817489;
RA Hammar M., Arqvist A., Bian Z., Olsen A., Normark S.;
RT "Expression of two csG operons is required for production of
RT fibronectin- and congo red-binding curli polymers in Escherichia coli
```

K-12.";
[1] Mol. Microbiol. 18:661-670(1995).
[2]
SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
[3]
SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikenoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kikagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA Res. 3:137-155(1996).
[4]
SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimallanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Ien G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
[5]
SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
[6]
SEQUENCE OF 1-21 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=95157246; PubMed=7854117;
RA Arngvist A., Olsen A., Normark S.;
RT "Sigma S-dependent growth-phase induction of the csfBA promoter in
Escherichia coli can be achieved in vivo by sigma 70 in the absence
of the nucleoid-associated protein H-NS.";
RL Mol. Microbiol. 13:1021-1032(1994).
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
CC CURLIN MONOMERS.
CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X90754; CAA62281.1; -.
DR
EMBL; D90741; BAA35831.1; -.
DR
EMBL; AE005315; AAG55787.1; -.
DR
EMBL; AP002554; BAB34842.1; -.
DR
PIR; C90806; C90806.
DR
PIR; G85665; G85665.
DR
PIR; S70787; S70787.
DR
Ecogene; EGI2821; csGB.
KW Fimbria; Signal; Complete proteome.
FT SIGNAL 1 21
FT CHAIN 22 151 MINOR CURLIN SUBUNIT.
SQ SEQUENCE 151 AA; 15882 MW; B18D266B964014B8 CRC64;
Query Match 13.6%; Score 104.5; DB 1; Length 151;
Best Local Similarity 29.1%; Pred No. 0.033; Indels 11; Gaps 3;
Matches 34; Conservative 15; Mismatches 57;
QY 38 SSGPSTLSIYQYGSANAALALQSDARKSETTITQSGYNGADVCGQADNSTIELTQNGF 97
Db 21 AAGYDLANSEYNF----AVNELSKSFNQAAIIGQAGTNSAQLRQGGSKLLAVVAQEGS 76
QY 98 RNNATYDQLVTRVVTHEMAH---AGNNAALVNQASDSSVMVROVGFNNATANQY 151
Db 77 SNRAKIDQ----TGDYNLAYIDQAGSANDASISQAGYGNVTAMIIQKGSNKANITQY 129
RESULT 5
CSGB_SALT1 STANDARD; PRT; 151 AA.
ID CSGB_SALT1
AC Q827M3; 2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Minor curlin subunit precursor.
GN CSGB OR STV1180 OR T1777.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
CX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrall B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodycanni V., Schwartz D.C., Blattner F.R.;
RT "Comparative genomics of Salmonella enterica serovar typhi strains Ty2
and CT18.";
RL J. Bacteriol. 185:2330-2337(2003).
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
CC CURLIN MONOMERS.
CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC	-----	
DR	EMBL; AL527269; CAD08257.1; -	
DR	EMBL; AE016840; AA069400.1; -	
KW	Fibrin; Signal; Complete proteome.	
FT	SIGNAL 1 21	POTENTIAL.
PT	CHAIN 22 151	MINOR CURLIN SUBUNIT.
SQ	SEQUENCE 151 AA; 16254 MW; 161C54326EE573495 CRC64;	

Query Match 13.0%; Score 99.5; DB 1; Length 151;
Best Local Similarity 28.2%; Pred. No. 0.088;
Matches 33; Conservative 16; Mismatches 51; Indels 17; Gaps 4;

```

QY      51  GSANPALALQSDARKSE-----TTTQSGYGVNGADVQO_GADNST-----IETQ  94
Db      14  GAGCIGATATNYDJARSEYFNAPVNELSKSFNQRAIIIGVGTGTDNSARVROEGSKLLSVISQ  73
QY      95  NGPRNNATYDQLVTRVVVTHEMAHAGGNNAALVQNTASDSSVMVRQVGFNNATANQY  151
Db      74  EGENNRKAVDQAGNYFNAY-IETGNANDASISQSAYGNSAAIIOKGSKNANITQY  129

```

RESULT 6

CSGB_SALTY	STANDARD;	PRT;	151 AA.
ID	CSGB_SALTY		
AC	P55226;		
DT	01-OCT-1996 (Rel. 34, Created)		
DT	01-OCT-1996 (Rel. 34, Last sequence update)		
DT	28-FEB-2003 (Rel. 41, Last annotation update)		
DE	Minor curlin subunit precursor (Fimbrin SEF17 minor subunit).		
GN	CSGB OR AGFB OR STM1143.		
OS	Salmonella typhimurium, and		
OS	Salmonella enteritidis.		
OC	Bacteria: Proteobacteria; Gammaproteobacteria; Enterobacteriales;		
OC	Enterobacteriaceae; Salmonella.		
OX	NCBI_TaxID=602, 592;		
	[1]		
RP	SEQUENCE FROM N.A.		
RC	SPECIES=S. typhimurium; STRAIN=SR-11;		
PX	MEDLINE=98117058; PubMed=9457850;		
RA	Reimling U., Bian Z., Hammar M., Sierralta W.D., Normark S.;		
RT	"Curli" fibers are highly conserved between <i>Salmonella typhimurium</i> and		
RT	<i>Escherichia coli</i> with respect to operon structure and regulation.";		
RL	J. Bacteriol. 180:722-731(1998).		

RP SEQUENCE FROM N.A.
RC SPECIES-S. Typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11877609;
RA McClelland M., Sanderson K.B., Spieth J., Clifton S.W., Latreille P.,
RACourney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvarey E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of *Salmonella enterica* serovar Typhimurium
LT2.";
RL Nature 413:852-856(2001).

RP SPECIES FROM N.A.
RC SPECIES ENTERITIDIS; STRAIN=27655-3B;
RD MEDLINE=56445512; PubMed=8550497;
RE Callinson S.K., Clouthier S.C., Doran J.L., Bansen P.A., Kay W.W.;
RF "Salmonella enteritidis agfBAC operon encoding thin, aggregative
RJ fimbriae."; RT
RL J. Bacteriol. 178:662-667(1996).

CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
CC CURLIN MONOMERS.

-|- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to licenses@isb-sib.ch).

```
CC -----
DR ENBL; AU002301; CAA05316.1; -.
DR ENBL; AE008749; AAL20073.1; -.
DR ENBL; U43280; AAC43598.1; -.
DR PIR; JC6040; JC6040.
DR styGene; SG10609; cscB.
```

KW	Fimbria; Signal; Complete proteome.
FT	POTENTIAL.
FT	MINOR CURLIN SUBUNIT.
SQ	SEQUENCE 151 AA; 16192 MW; C0FC5430E6DD361D CRC64;

Query Match	13.0%	Score 99.5	DB 1	Length 151
Best Local Similarity	28.2%	Pred. No. 0.088		
Matches 33	Conservative 16	Mismatches 51	Indels 17	Caps 4

```
QY      51 GSANAALAQSDARKSE-----TTTQTSGYNGADVGO-GADNST-----HETLQ 94
       | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
       | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB     14 GAPGATATNDYLARSEYNFAVNEKSSFNQAIIIGQGTGDTNSARVRQEGSKLLSVISQ 73
       | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

QY 95 NGFRNNATYDCLVTRVVTHEMAHAGGNNALVNQTRASDSSVWVRQVGFGNNATANQY 151
:
74 EGGNNRAKVPQAGNYNFAY-IEQTGNANDASISQSAYGNSAAIIQRXSGNKANITQY 129
:

RESULT 7 OVO DROME

AC	P51521; Q9XZU4;	(Rel. 34)
DT	01-OCT-1996	(Rel. 34, Created)
DT	01-OCT-1996	(Rel. 34, Last sequence update)

DE OVO protein (Shaven baby protein).
DE OVO CR. SVE.
OS Drosophila melanogaster (Fruit fly).

CC Caneblow, Hymenoptera, Hymenoptera, insecta, Pterygota;
CC Neopetia, Enderogaster, Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI TaxID=7227

RP SEQUENCE FROM N.A.
 TISSUE=Ovary;
 RC
 RX MEDLINE=95021309; PubMed=7935398;

RT Multiple products from the shavenb-cvo gene region of *Drosophila*
RT melanogaster: relationship to genetic complexity.";
RL Mol. Cell. Biol. 14: 6809-6818(1994).

RP SEQUENCE FROM N.A.
RC STRAIN=Oregon-K;
RX MEDLINE=91293102; PubMed=1712294;

"The ovo gene of *Drosophila* encodes a zinc finger protein required for female germ line development.";
RL EMBO J. 10:2259-2266 (1991).

CC LINE CELLS. PLAYS A ROLE IN GERM LINE SEX DETERMINATION.
CC - SUBCELLULAR LOCATION: Nuclear (Potential).
CC - DEVELOPMENTAL STAGE: FIRST APPEARS IN THE GERMARIIUM AND

CC BUT IS RAPIDLY LOST IN THE EMBRYOS EXCEPT FOR ITS CONTINUED
CC PRESENCE IN THE GERM LINE PRECURSOR POLE CELLS.
CC - SIMILARITY: Contains 4 C2H2-type zinc fingers.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

Tankyrase 1 (EC 2.4.2.30) (TANK1) (Tankyrase 1) (TNKS-1) (TRF1-interacting ankyrin-related ADP-ribose polymerase).
TNKS OR TNKS1 OR TINI OR TINI1 OR PARPL.
Homo sapiens (Human).
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;

EMBL; U11383; AAB60216.1; -
PIR; A56038; CAB36921.1; ALT_SEQ.
HSSP; P07248; 2ADR.
TRANSFAC; T006569; -
Flybase; FBgn003028; ovo.
InterPro; IPR007087; Znf_C2H2.
Pfam; PF00096; zf-C2H2; 3.
SMART; SM00355; Znf_C2H2; 4.
PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
PROSITE; PS00157; ZINC_FINGER_C2H2_2; 3.
Zinc-finger; Metal-binding; DNA-binding; Repeat; Nuclear protein;
Transcription regulation.
DOMAIN 62 66 POLY-ALA.
DOMAIN 72 77 POLY-GLY.
DOMAIN 80 85 POLY-GLY.
DOMAIN 98 108 POLY-GLY.
DOMAIN 144 152 POLY-HIS.
DOMAIN 153 159 POLY-ASN.
DOMAIN 153 339 POLY-GLN.
DOMAIN 336 353 POLY-GLN.
DOMAIN 347 353 POLY-GLN.
DOMAIN 357 361 POLY-GLN.
DOMAIN 410 414 POLY-GLN.
DOMAIN 422 422 POLY-GLN.
DOMAIN 426 432 POLY-GLN.
DOMAIN 426 432 POLY-GLN.
DOMAIN 435 453 POLY-GLN.
DOMAIN 456 459 POLY-GLN.
DOMAIN 466 474 POLY-ALA.
DOMAIN 474 517 POLY-ALA.
DOMAIN 497 517 POLY-SER.
DOMAIN 524 529 POLY-ALA.
DOMAIN 549 558 POLY-ALA.
DOMAIN 639 651 POLY-ALA.
DOMAIN 717 725 POLY-GLN.
DOMAIN 797 802 POLY-GLN.
DOMAIN 820 823 POLY-GLN.
DOMAIN 826 832 POLY-GLN.
DOMAIN 874 896 C2H2-TYPE 1.
DOMAIN 902 924 C2H2-TYPE 2.
DOMAIN 930 953 C2H2-TYPE 3.
DOMAIN 959 982 C2H2-TYPE 4.
DOMAIN 647 647 A -> R (IN REF. 2).
SEQUENCE 1028 AA; D7068BB2BC0F677 CRC64;

Query Match 12.0%; Score 92.5; DB 1; Length 1028;
Best Local Similarity 26.0%; Pred. No. 2.9;
Matches 38; Conservative 15; Mismatches 66; Indels 27; Gaps 4;
QY 3 LKVRFAAIVVSGALAGVQWGGGNGHNGGNSGPDSTLSIYQYGSANAALQSD 62
Db 59 LQNAAAAYVINGAGG-----GGCTGGGAGSGPFGGFSANSGGGGGG----- 104
QY 63 ARKSETTITQSGYNGADYGGADNSGTIELTNGFRNNATYDQLVTRVVTHEMAHAGNN 122
Db 105 -----GGNGYINGGVG-CPNNSLDGNLLNFASVSNYESNKFHHHHHHQHNNN 155
QY 123 AALVNOTASDSSVMVRQVFGNNATA 148
Db 156 ----NNNNGGQTSMMGHPYGGNPSA 177

RESULT 8
TNK1_HUMAN
ID TNK1_HUMAN STANDARD; PRT; 1327 AA.
AC O95271; O95272;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Tankyrase 1 (EC 2.4.2.30) (TANK1) (Tankyrase 1) (TNKS-1) (TRF1-interacting ankyrin-related ADP-ribose polymerase).
DE TNKS OR TNKS1 OR TINI OR TINI1 OR PARPL.
DE OS Homo sapiens (Human).
DE OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
DE OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
DE OX NCBI_TaxID=9606;
DE RN [1]
DE RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
DE RC TISSUE=Testis;
DE RX MEDLINE=99040105; PubMed=9822378;
DE RA Smith S., Giriat I., Schmitt A., de Lange T.;
DE RT "Tankyrase, a poly(ADP-ribose) polymerase at human telomeres.";
DE RL Science 282:1484-1487(1998).
DE RN [2]
DE RP SUBCELLULAR LOCATION.
DE RX MEDLINE=99454782; PubMed=10523501;
DE RA Smith S., de Lange T.;
DE RT "Cell cycle dependent localization of the telomeric PARP, tankyrase,
DE RT to nuclear pore complexes and centrosomes.";
DE RL J. Cell Sci. 112:3649-3656(1999).
DE RN [3]
DE RP FUNCTION, AND PHOSPHORYLATION.
DE RX MEDLINE=20556282; PubMed=10988299;
DE RA Chi N.-W., Lodish H.F.;
DE RT "Tankyrase is a Golgi-associated mitogen-activated protein kinase
DE RT substrate that interacts with IRAP in GLUT4 vesicles.";
DE RL J. Biol. Chem. 275:38437-38444(2000).
DE RN [4]
DE RP FUNCTION, AND MUTAGENESIS OF HIS-1184 AND GLU-1291.
DE RX MEDLINE=21602874; PubMed=11739745;
DE RA Cook B.D., Dyrnek J.N., Chang W., Shostak G., Smith S.;
DE RT "Role for the related poly(ADP-Ribose) polymerases tankyrase 1 and 2
DE RT at human telomeres.";
DE RL Mol. Cell. Biol. 22:332-342(2002).
DE CC -!- FUNCTION: May regulate vesicle trafficking and modulate the
DE CC subcellular distribution of SLC2A4/GLUT4-vesicles. Has PARP
DE CC activity and can modify TRF1, and thereby contribute to the
DE CC regulation of telomere length.
DE CC -!- CATALYTIC ACTIVITY: NAD(+) + (ADP-D-riboseyl)(N)-acceptor =
DE CC nicotinamide + (ADP-D-riboseyl)(N+1)-acceptor.
DE CC -!- SUBUNIT: Oligomerizes and associates with TNKS2. Interacts with
DE CC the cytoplasmic domain of LNPBP/Oase in SLC2A4/GLUT4-vesicles.
DE CC Binds to the N-terminus of telomeric TRF1 via the ANK repeats.
DE CC -!- SUBCELLULAR LOCATION: Cytoplasmic, associated with the Golgi and
DE CC with juxtanuclear SLC2A4/GLUT4-vesicles. A minor proportion is
DE CC also found at nuclear pore complexes and around the pericentriolar
DE CC matrix of mitotic centrosomes. During interphase, a small fraction
DE CC of TNKS is found in the nucleus, associated with TRF1.
DE CC -!- ALTERNATIVE PRODUCTS:
DE CC Event=Alternative splicing; Named isoforms=2;
DE CC Name=1;
DE CC IsoId=O95271-1; Sequence=Displayed;
DE CC Name=2;
DE CC IsoId=O95271-2; Sequence=VSP_004538, VSP_004539;
DE CC Note=No experimental confirmation available;
DE CC -!- TISSUE SPECIFICITY: Ubiquitous; highest levels in testis.
DE CC -!- PTM: Upon insulin-stimulation, phosphorylated on serine residues
DE CC by MAPK kinases.
DE CC -!- PTM: ADP-ribosylated (-auto).
DE CC -!- SIMILARITY: Belongs to the PARP family.
DE CC -!- SIMILARITY: Contains 15 ANK repeats.
DE CC -!- SIMILARITY: Contains 1 sterile alpha motif (SAM) domain.
DE CC This SWISS-PROT entry is copyright. It is produced through a collaboration
DE CC between the Swiss Institute of Bioinformatics and the EMBL outstation
DE CC the European Bioinformatics Institute. There are no restrictions on its
DE CC use by non-profit institutions as long as its content is in no way
DE CC modified and this statement is not removed. Usage by and for commercial
DE CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
DE CC or send an email to license@isb-sib.ch).
DE CC -----
DE CC EMBL; AF082556; AAC79841.1; -

```

DR EMBL; AF082557; AAC79842.1; -.
DR EMBL; AF082558; AAC79843.1; -.
DR EMBL; AF082559; AAC79844.1; -.
DR HSSP; Q00420; LAMC.
DR MIM; 603303; -.
DR GO; GO:0000781; C:chromosome, telomeric region; IDA.
DR GO; GO:0003950; F:NAD ADP-ribosyltransferase activity; IDA.
DR GO; GO:0005515; F:protein binding; IPI.
DR GO; GO:0007004; P:telomerase-dependent telomere maintenance; NAS.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR001660; SAM.
DR Pfam; PF00023; ank; 19.
DR Pfam; PF00536; SAM; 1.
DR PRINTS; PR01415; ANKYRIN.
DR SMART; SM00248; ANK; 17.
DR SMART; SM00454; SAM; 1.
DR PROSITE; PS50088; ANK_REPEAT; 15.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS50105; SAM_DOMAIN; 1.
KW Transferase; Glycosyltransferase; NAD; Golgi stack; Telomere;
KW Nuclear protein; Repeat; ANK repeat; ADP-ribosylation;
KW Phosphorylation; Alternative splicing.
FT REPEAT 215 247
FT REPEAT 248 280
FT REPEAT 281 313
FT REPEAT 368 400
FT REPEAT 401 433
FT REPEAT 434 466
FT REPEAT 521 556
FT REPEAT 557 589
FT REPEAT 590 622
FT REPEAT 683 715
FT REPEAT 716 748
FT REPEAT 749 781
FT REPEAT 836 868
FT REPEAT 869 901
FT REPEAT 902 934
FT DOMAIN 1030 1089
FT DOMAIN 1176 1327
FT DOMAIN 9 14
FT DOMAIN 27 34
FT DOMAIN 128 134
FT DOMAIN 137 145
FT VARSPLIC 641 643
FT VARSPLIC 544 1327
FT MUTAGEN 1184 1184
FT MUTAGEN 1291 1291
FT MUTAGEN 1291 1291
FT SEQUENCE 1327 AA; 142010 MW; E14DE985C710B957 CRC64;

Query Match 11.9%; Score 91.5; DB 1; Length 1327;
Best Local Similarity 29.4%; Pred. No. 4.7;
Matches 42; Conservative 17; Mismatches 65; Indels 19; Gaps 6;

QY 6 VAAPAAI-VVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQVGSANAALQSDAR 64
Db 99 VAAPVAVNSTSAAAGVAVNPAGSGNNPSSSSPTSS-SSSSPSSGSSLAESPEAA 157

QY 65 KSETTIT-OSQYNGADVGCGADNSTIETQNGFRNNATYDQLVTRVVTHEMAHAGG 120
Db 158 GVSSTAPLPGAGGPGTGVPAVSGALRELLACRNGD-----VSRV--KRLVDAAN 206

QY 121 NNAALVNQTSADSSVMVQVGF 143
Db 207 VNAK--DMAGKSSPLHFAAGFG 227

RESULT 9
OMPB_RICJA

ID OMPB_RICJA STANDARD; PRT; 1656 AA.
AC O06653;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Outer membrane protein B precursor (188 kDa surface-layer protein)
DE (Surface protein antigen) (Cell surface antigen 5) (Scas) (rOmpB)
DE (rOmp B) (Contains: 120 kDa surface-exposed protein (Surface protein
DE antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide).
GN OMPB.
OS Rickettsia japonica.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=35790;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YH;
RA Uchiyama T.;
RT Sequencing of the gene encoding the protein rOmp B of Rickettsia
RT japonica.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.
CC -!- FUNCTION: THE 120 kDa SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (By
CC similarity).
CC -!- FUNCTION: THE 32 kDa BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR
CC (By similarity).
CC -!- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-
CC layer with hexagonal symmetry.
CC -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB003681; BAA20138.1; -.
CC InterPro; IPR006315; Autotransport.
CC InterPro; IPR005546; Autotransporter.
CC Pfam; PF03797; Autotransporter; 1.
CC TIGRFAMs; TIGR01414; autotrans_bar1; 2.
KW Antigen; S-layer; Cell wall.
FT CHAIN 1 1338 120 kDa SURFACE-EXPOSED PROTEIN.
FT CHAIN 1339 1656 32 kDa BETA PEPTIDE.
FT DOMAIN 528 533 POLY-GLY.
FT SEQUENCE 1656 AA; 168097 MW; 3132A69C9DD5999F CRC64;

Query Match 11.7%; Score 90; DB 1; Length 1656;
Best Local Similarity 23.5%; Pred. No. 8;
Matches 40; Conservative 21; Mismatches 57; Indels 52; Gaps 6;

QY 6 VAAPAAI-VVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQVGSANAALQSDARK 65
Db 509 VLAAGAITLDGSAIT-----TGDIGNGG-----GAALQSITLANDATK 547

QY 66 SETTITQSG-----YNGADVGCGADNSTIET-----QNGFRNNATYD 104
Db 548 ----TTLGGANIISANGGTINFQANGGTIKTSTNNIVDCDLAIAITDQTGVDPASSLT 604

QY 105 QLTVRVVTHEMAHAGGNAAL-----VNQTSADSSVMVQVGFNNAT 147
Db 605 NQAQTLTISGTIGIICANNITLGGQFNIGSKTTLNGVAINELVIGNNS 654

RESULT 10
N100 YEAST
ID N100 YEAST STANDARD; PRT; 959 AA.
AC Q02629;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)

```



```

DR EMBL; D89145; BAAL3807.1; -.
DR PIR; T50074; T50074.
DR HSSP; P22629; 1SWF.
DR GeneDB SPombe; SPAC1486.05; -.
DR GO; GO:0016020; C:membrane; ISS.
DR GO; GO:0005643; C:nuclear pore; ISS.
DR GO; GO:0006606; P:protein-nucleus import; ISS.
DR InterPro; IPR007230; Nucleoporin2.
DR InterPro; IPR004325; Nucleoporin_FG.
DR Pfam; PF04096; Nucleoporin2; 1.
DR Pfam; PF03093; Nucleoporin_FG; 24.
KW Nuclear protein; Transport; Repeat.
FT DOMAIN 9 625 GLY-RICH.
FT DOMAIN 233 596 THR-RICH.
FT DOMAIN 393 442 ASN-RICH.
SQ SEQUENCE 1778 AA; 189575 MW; 1D2AED57D927ADC6 CRC64;

Query Match 11.5%; Score 88; DB 1; Length 1778;
Best Local Similarity 27.3%; Pred. No. 13;
Matches 38; Conservative 16; Mismatches 65; Indels 20; Gaps 5;

QY 15 SGSALAGYVFPWGGGNGHNGSGSPDSTL-----SIYQGSANAALALQSDARKSETT 69
Dd : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
400 SGTWGTGL--FGGANNTANNNTAPTSTFGGNSNFSGANNAATKPSGFGSGTT 456

QY 70 ITQSGYGNADVGGADNSTIELTQNGFRNNAITDQLVTRVVRVTHEMAHAGNNALVNOT 129
Dd : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
457 TTPA--SGGFGGQANNA-----PKPAFGSTATTAPKPAFTGLFGGLGAGANTTATNAT 510

QY 130 ASDSSVMVRQVQFGGNATA 148
Dd : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
511 GTGQSL-----FGNANTA 523

RESULT 12
VG38 BPT2
ID -VG38 BPT2 \ STANDARD; PRT; 262 AA.
AC P07875;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DE Receptor recognizing protein (Protein Gp38).
GN 38.
OS Bacteriophage T2.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
OC T4-like viruses.
OC NCSI_TaxID=10664;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87283911; PubMed=3302276;
RA Riede I., Drexler K., Eschbach M.L., Henning U.;
RT "DNA sequence of genes 38 encoding a receptor-recognizing protein of
bacteriophages T2, X3 and of X3 host range mutants.";
RL J. Mol. Biol. 194:31-39(1987).
CC -!- FUNCTION: VG38 is at the tip of the long tail fibers and serves as
the phage recognition site for the cellular receptor.
CC -!- MISCELLANEOUS: THIS PHAGE USE OUTER MEMBRANE PROTEINS OMPF AND TTR
AS RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X05312; CAA28935.1; -.
DR PIR; S00275; S00275.
DR InterPro; IPR007932; Tail_fibre_GP38.
DR Pfam; PF05268; GP38; 1.
KW Fiber protein; Phage recognition.
SQ SEQUENCE 262 AA; 25801 MW; 0567366918F6C745 CRC64;

Query Match 11.3%; Score 86.5; DB 1; Length 262;
Best Local Similarity 44.8%; Pred. No. 2;
Matches 26; Conservative 4; Mismatches 19; Indels 9; Gaps 3;

QY 27 GCGGNHNGGNSGPDSTLSIYQGSANAALALQSDARKSETTITQSGYGNADVGG 84
Dd : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
175 GGGRRPFGVGKIGSDSILS-----GSNASL---TDAGTGGTTF-QYGAGNGGNVGAG 223

RESULT 13
GP63 LEIDO
ID -GP63 LEIDO STANDARD; PRT; 590 AA.
AC P23223;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Leishmanolysin precursor (EC 3.4.24.36) (Cell surface protease)
DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface
endopeptidase).
GN GP63.
OS Leishmania donovani.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OC NCSI_TaxID=5661;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LV9;
RX MEDLINE=92107220; PubMed=1762629;
RA Webb J.R., Button L.L., McMaster R.W.;
RT "Heterogeneity of the genes encoding the major surface glycoprotein
of Leishmania donovani";
RL Mol. Biochem. Parasitol. 48:173-184(1991).
CC -!- FUNCTION: Has an integral role during the infection of macrophages
in the mammalian host.
CC -!- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and
P1' and basic residues at P2 and P3'. A model nonapeptide is
cleaved at -Ala-Tyr-I-Leu-Lys-Lys-.
CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity)
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -!- SIMILARITY: Belongs to peptidase family M8.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M60048; AAA29244.1; -.
DR HSSP; P08148; 1LML.
DR MEROPS; M08.001;
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR InterPro; IPR001577; Peptidase_M8.
DR Pfam; PF01457; Peptidase M8; 1.
DR PRINTS; PR00782; LSHMANOLYSIN.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydrolase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;
Zymogen; Signal; Cell adhesion; GPI-anchor; Lipoprotein.
FT SIGNAL 1 39 POTENTIAL.
FT PROPEP 40 87 ACTIVATION PEPTIDE.
FT CHAIN 88 565 LEISHMANOLYSIN.
FT PROPEP 566 590 REMOVED IN MATURE FORM (BY SIMILARITY).
FT METAL 251 251 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 252 252 BY SIMILARITY.
FT METAL 255 255 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 321 321 ZINC (CATALYTIC) (BY SIMILARITY).
FT DISULFID 112 129 BY SIMILARITY.
FT DISULFID 178 217 BY SIMILARITY.
FT DISULFID 301 373 BY SIMILARITY.
FT DISULFID 380 443 BY SIMILARITY.
FT DISULFID 393 412 BY SIMILARITY.
FT DISULFID 402 477 BY SIMILARITY.

```

```

FT DISULFID 454 498 BY SIMILARITY.
FT DISULFID 503 553 BY SIMILARITY.
FT DISULFID 523 546 BY SIMILARITY.
FT CARBOHYD 287 287 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT LIPID 565 565 GPI-anchor amidated asparagine (By similarity).
SQ SEQUENCE 590 AA; 62950 MW; 0PB315D299659F58 CRC64;

Query Match 11.3%; Score 86.5; DB 1; Length 590;
Best Local Similarity 47.1%; Pred. No. 4.9;
Matches 24; Conservative 4; Mismatches 18; Indels 5; Gaps 2;

QY 103 YDQLVTRVVTHMAHAGNNAALVQNTASDSVM-VROVGFC---NNATA 148
      |||||
Db 241 YDQLVTRVVTHMAHAGLGSVVFPRDARILEISNVRHKDFDPVINSSTA 291
      |||||

RESULT 14
GP63 LEIME STANDARD; PRT; 646 AA.
AC P43150;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Leishmanolysin C1 precursor (EC 3.4.24.36) (Cell surface protease)
DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface endopeptidase).
DE GP63-C1.
GN GP63-C1.
OS Leishmania mexicana.
OC Eukaryota; Eulenczoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5665;
RN [1]
RP RC SEQUENCE FROM N.A.
RC STRAIN=MNYC/BZ/62/M379;
RX MEDLINE=93149206; PubMed=8426614;
RA Medina-Acosta E., Kares R.E., Russell D.G.;
RT "Structurally distinct genes for the surface protease of Leishmania mexicana are developmentally regulated.";
RL Mol. Biochem. Parasitol. 57:31-46(1993).
CC -!- FUNCTION: Has an integral role during the infection of macrophages in the mammalian host.
CC -!- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and P1', and basic residues at P2 and P3'. A model nonapeptide is cleaved at -Ala-Tyr-|-Leu-Lys-Lys-.
CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC -!- DEVELOPMENTAL STAGE: Expressed in both the promastigote and the amastigote forms.
CC -!- SIMILARITY: Belongs to peptidase family M8.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC
CC EMBL; X64394; CAA45733.1; -.
CC PIR; S19916; S19916.
CC HSSP; P08148; 1LML.
CC MEROPS; M08.001; -.
CC GlycoSuiteDB; P43150; -.
CC InterPro; IPR006025; Pept_M_Zn_BS.
CC InterPro; IPR001577; Peptidase_M8.
CC Pfam; PF01457; Peptidase_M8_1.
CC PRINTS; PR00782; LSHMANOLYSIN.
CC PROSITE; PS00142; ZINC_PROTEASE; 1.
CC Hydrolase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;
KW Zymogen; Signal; Cell adhesion; Multigene family.
FT SIGNAL 1 39 POTENTIAL.
FT PROPEP 40 102 ACTIVATION PEPTIDE (POTENTIAL).
FT CHAIN 103 646 LEISHMANOLYSIN C1.
FT METAL 266 266 ZINC (CATALYTIC) (BY SIMILARITY).

```

```

FT ACT_SITE 267 267 BY SIMILARITY.
FT METAL 270 270 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 336 336 ZINC (CATALYTIC) (BY SIMILARITY).
FT DISULFID 127 144 BY SIMILARITY.
FT DISULFID 193 232 BY SIMILARITY.
FT DISULFID 316 388 BY SIMILARITY.
FT DISULFID 395 458 BY SIMILARITY.
FT DISULFID 408 427 BY SIMILARITY.
FT DISULFID 417 492 BY SIMILARITY.
FT DISULFID 469 513 BY SIMILARITY.
FT DISULFID 518 568 BY SIMILARITY.
FT DISULFID 538 581 BY SIMILARITY.
FT CARBOHYD 86 86 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 297 297 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 399 399 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 409 409 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 433 433 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 445 445 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 466 466 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 501 501 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 646 AA; 69054 MW; FE48DDC78C10B0A CRC64;

Query Match 11.2%; Score 86; DB 1; Length 646;
Best Local Similarity 94.4%; Pred. No. 6;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 103 YDQLVTRVVTHMAHAGG 120
      |||||
Db 256 YDQLVTRVVTHMAHAGV 273
      |||||

RESULT 15
ICEN_XANCT STANDARD; PRT; 1567 AA.
AC P18127;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ice nucleation protein.
GN INAX.
OS Xanthomonas campestris (pv. translucens).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=343;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=X56S;
RX MEDLINE=91080859; PubMed=2259339;
RA Zhao J., Orser C.S.;
RT "Conserved repetition in the ice nucleation gene inax from Xanthomonas campestris pv. translucens.";
RL Mol. Gen. Genet. 223:163-166(1990).
CC -!- FUNCTION: Ice nucleation proteins enable bacteria to nucleate crystallization in supercooled water. (By similarity).
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- DOMAIN: CONTAINS 153 IMPERFECT REPEATS OF THE CONSENSUS OCTAPEPTIDE A-G-Y-G-S-T-L-T; FURTHER ON A 16-RESIDUE AND A REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.
CC -!- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.
CC -!- SIMILARITY: Belongs to the bacterial ice nucleation protein family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC
CC EMBL; X52970; CAA37140.1; -.
CC HSSP; P06620; 1INA.

```

Search completed: August 2, 2004, 14:49:33
Job time : 6.3 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:39:17 ; Search time 29.7 Seconds

(without alignments)
1604.150 Million cell updates/sec

Title: US-09-543-407-30

Perfect score: 768

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVQVGFNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phase:*

10: sp_plant:*

11: sp_rodent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_rvirus:*

16: sp_bacteriap:*

17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	672	87.5	152	2	O33802		O33802 salmonella
2	570.5	74.3	150	2	O7X243		O7X243 citrobacter
3	534	69.5	149	2	O7X240		O7X240 citrobacter
4	506.5	66.0	152	16	Q8CW63		Q8CW63 escherichia
5	435.5	56.7	150	2	Q7X237		Q7X237 enterobacte
6	380	49.5	176	2	Q34069		Q34069 salmonella
7	126	16.4	502	16	Q8E1H4		Q8E1H4 shewanella
8	122	15.9	29	2	Q8S3J5		Q8S3J5 escherichia
9	112	14.6	139	16	Q8E1H3		Q8E1H3 shewanella
10	111.5	14.5	171	16	Q89J13		Q89J13 bradyrhizob
11	105	13.7	130	16	Q89J14		Q89J14 bradyrhizob
12	104.5	13.6	151	16	Q7UCZ1		Q7UCZ1 shigella fl
13	104.5	13.6	160	16	Q8CW64		Q8CW64 escherichia
14	104.5	13.6	160	16	Q83RU7		Q83RU7 shigella fl
15	104	13.5	157	16	Q88HG0		Q88HG0 pseudomonas
16	102.5	13.3	3659	16	Q98LN6		Q98LN6 rhizobium 1

17	102	13.3	1422	16	Q8EFU3	Q8efu3 shewanella
18	101.5	13.2	531	13	Q8AWA8	Q8awa8 lampetra fl
19	101	13.2	1748	5	Q94821	Q94821 tetrahymena
20	101	13.2	7716	16	Q7UWZ8	Q7uwz8 rhodopirell
21	99.5	13.0	1209	16	Q89CK5	Q89ck5 bradyrhizob
22	99.5	13.0	2174	16	Q82U08	Q82u08 rhizobium m
23	99	12.9	562	16	Q89D03	Q89d03 bradyrhizob
24	99	12.9	441	2	O87327	O87327 pseudomonas
25	98.5	12.8	348	13	Q93397	Q93397 cyprinus ca
26	98.5	12.8	624	3	Q8N1V1	Q8n1v1 neurospora
27	98	12.8	145	16	Q8U6N9	Q8u6n9 agrobacteri
28	97.5	12.7	91	2	Q9S3J8	Q9s3j8 escherichia
29	97.5	12.7	477	2	Q7X4S5	Q7x4s5 pseudomonas
30	97.5	12.7	586	16	Q8EXJ2	Q8exj2 leptospira
31	97.5	12.7	845	16	Q7U1C5	Q7u1c5 mycobacteri
32	97.5	12.7	646	16	O53818	O53818 mycobacteri
33	97	12.6	1410	16	O8CMJ0	O8cmj0 shewanella
34	96.5	12.6	191	10	Q7XDR3	Q7xdr3 oryza sativ
35	96.5	12.6	313	10	Q9SYZ5	Q9syz5 arabidopsis
36	95.5	12.4	151	2	O7X244	O7x244 citrobacter
37	95.5	12.4	154	16	Q89J15	Q89j15 bradyrhizob
38	95.5	12.4	346	5	Q9SR55	Q9srs5 drosophila
39	95.5	12.4	908	5	Q9VM71	Q9vm71 drosophila
40	95	12.4	151	2	O7X238	O7x238 enterobacte
41	95	12.4	196	10	O22638	O22638 zea mays (m
42	94.5	12.3	153	16	O89J16	O89j16 bradyrhizob
43	94.5	12.3	908	5	Q9VEC4	Q9vec4 drosophila
44	94.5	12.3	909	5	Q8T4E0	Q8t4e0 drosophila
45	94.5	12.3	1713	3	Q8TGE1	Q8tge1 saccharomyc

ALIGNMENTS

RESULT 1

O33802 PRELIMINARY; PRT; 152 AA.
ID O33802;
AC O33802;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE AgfA protein (Fragment).
GN AGFA.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
CX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE:980533981; PubMed:9393832;
RA Sukupolvi S.S., Lorentz R.G., Gordon J.I., Bian Z., Pfeifer J.D.,
RA Normark S.J., Rhen M.;
RT "Expression of thin, aggregative fimbriae promotes interaction of
RT Salmonella typhimurium SR-11 with mouse small intestinal epithelial
RT cells".
RL Infect. Immun. 65:5320-5325 (1997).
DR EMBL; AJ000514; CA004151.1;
FT NON_TER 152 152
SQ SEQUENCE 152 AA; 15401 MW; 9DA7DADC2364B006 CRC64;

Query Match 87.5%; Score 672; DB 2; Length 152;

Best Local Similarity 89.4%; Pred. No. 38-46;

Matches 135; Conservative 3; Mismatches 13; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSSGPDSTLSIYQYGSANAALALQ 60

Db 1 MKLLKVAFAAIVVSGSAGVVPQWGGGNGGNGSSGPDSTLSIYQYGSANAALALQ 60

QY 61 SPARKSETTITOSGYNGADVCGGADNSTIETQNGFRNNATYDQLVTWVTHEVAHAGG 120

Db 61 SPARKSETTITOSGYNGADVCGGADNSTIETQNGFRNNATYDQLVTWVTHEVAHAGG 120

QY 121 NNAALVNQTASDSSVMVQVGFNNATANQY 151


```
RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RT Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158 (2003).
DR EMBL; AJ515702; CAD56678.1; -.
SQ SEQUENCE 150 AA; 15112 MW; 5D8BB2D872DF15F3 CRC64;

Query Match 56.7%; Score 435.5; DB 2; Length 150;
Best Local Similarity 60.3%; Pred. No. 2.2e-27;
Matches 91; Conservative 26; Mismatches 33; Indels 1; Gaps 1;

Qy 1 MKLLKVAFAAIVVSGSALAGVVPQMGNGNHNHGGNSGPDSTLSIYQYGSANAALALQ 60
Db 1 MKPKVAAALAAIVVSGSAGAMINQ-CGWGHGCHGGYGGPNSTLNIYQNGGNSALALQ 59

Qy 61 SDARKSETTTTQSGYNGADVGQADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAG 120
Db 60 TDARNSVLNISTQGGNGADVGQSDSSINLTQNGFGNSATLDQWNSKDSVMNVQSYGG 119

Qy 121 NNAALVNOTASDSSVMVQVQFGNNATANOV 151
Db 120 LNALVDQTASNSTVNTVQTGFGNHATAHQV 150

RESULT 6
Q54069 PRELIMINARY; PRT; 76 AA.
AC Q54069,
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SEF17 fimbria (Fragment).
GN AGFA.
OS Salmonella enteritidis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=592;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SE30;
RA Cox J.M., Eglezos S., Woolcock J.B.;
RT "Virulence of Salmonella enteritidis in chickens correlates with
RT colony morphology and expression of SEF17 fimbriae.";
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U53207; AAA98671.1; -.
FT NON_TER 1
FT NON_TER 76
SQ SEQUENCE 76 AA; 7704 MW; 2FD5411241A7BCB1 CRC64;

Query Match 49.5%; Score 380; DB 2; Length 76;
Best Local Similarity 97.4%; Pred. No. 2.7e-23;
Matches 74; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 30 GNHNGGNSGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGADVGQADNST 89
Db 1 GNHNGGNSGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGADVGQADNST 60

Qy 90 IELTQNGFRNNATYDQ 105
Db 61 IELTQNGFRNNATIDQ 76

RESULT 7
Q8EIH4 PRELIMINARY; PRT; 502 AA.
AC Q8EIH4,
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Conserved hypothetical protein.
GN S00865.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.
```

```
OX NCBI_TaxID=70863;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MR-1;
RX MEDLINE=22297686; PubMed=12368813;
RA Heideberg J.F., Paulsen I.T., Nelson K.B., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
RA Vamathevan J., Weidman J., Imbraim M., Lee K., Berry K., Lee C.,
RA Mueller J., Khouri H., Gill J., Uterback T.R., McDonald L.A., C.M.;
RA Fiedlblyum T.V., Smith H.O., Venter J.C., Nealsen K.H., Fraser B.;
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis.";
RL Nat. Biotechnol. 20:1118-1123 (2002).
DR EMBL; AF015532; AAN53941.1; -.
DR TIGR; S00865; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 502 AA; 52441 MW; D08CA23D6C46B62D CRC64;

Query Match 16.4%; Score 126; DB 16; Length 502;
Best Local Similarity 27.0%; Pred. No. 0.045;
Matches 41; Conservative 22; Mismatches 57; Indels 32; Gaps 5;

Qy 29 GGNHNG-----GGN-----SSGPDSTLSIYQYGSANA---ALALQS 61
Db 231 GDNHTGFVVALAGSENDISMEQEGSNNTAYLSMTTGDDNTVDITQDGSNTVYDLSIADI 290

Qy 62 DARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGN 121
Db 291 QGDDNDITIKQKGDNGAEFQWGDSDVDLQKQGDANFATFGAYGTD-NDFDLSSKGDN 349

Qy 122 NNAALVNOTASDSSVMVQVQFGN-----NATAN 149
Db 350 NELVAFATGEDNSIEISQEGDANFAYVDATGN 381

RESULT 8
Q9S3J5 PRELIMINARY; PRT; 29 AA.
AC Q9S3J5,
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE Curlin subunit monomer (Fragment).
GN CSGA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC TRANSPOSON-Insertion sequence IS1;
RX MEDLINE=99314153; PubMed=10386375;
RA La Ragione R.M., Collighan R.J., Woodward M.J.;
RT "Non-curlin of Escherichia coli O78:k80 isolates associated with
RT IS1 insert on in csGB and reduced persistence in poultry infection.";
RL FEMS Microbiol. Lett. 175:247-253 (1999).
DR EMBL; AJ131756; CAB45380.1; -.
FT NON_TER 29
FT NON_TER 29
SQ SEQUENCE 29 AA; 2789 MW; E290DFC07ABBB243 CRC64;

Query Match 15.9%; Score 122; DB 2; Length 29;
Best Local Similarity 89.7%; Pred. No. 0.0035;
Matches 26; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MKLLKVAFAAIVVSGSALAGVVPQMGNG 29
Db 1 MKLLKVAFAAIVVSGSALAGVVPQMGNG 29

RESULT 9
```

Q8EIH3	Q8EIH3	PRELIMINARY;	PRT;	139 AA.	
ID	Q8EIH3	PRELIMINARY;	PRT;	139 AA.	
AC	Q8EIH3	PRELIMINARY;	PRT;	139 AA.	
DT	01-MAR-2003 (Tremblrel. 23, Created)				
DT	01-MAR-2003 (Tremblrel. 23, Last sequence update)				
DT	01-MAR-2003 (Tremblrel. 23, Last annotation update)				
DE	Minor curlin subunit CsgB, putative.				
GN	SO0866.				
OS	Shewanella oneidensis.				
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;				
OC	Alteromonadaceae; Shewanella.				
OC	NCBI_TaxID=70863;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=NR-1;				
RX	MEDLINE=22297686; PubMed=12368913;				
RA	Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,				
RA	Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,				
RA	Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,				
RA	DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,				
RA	Madupu R., Peterson J.D., Unayam L.A., White C., Wolf A.M.,				
RA	Vamathevan J., Weidman J., Imprial M., Lee K., Berry K., Lee C.,				
RA	Mueller J., Khouri H., Gill J., Uterback T.R., McDonald L.A.,				
RA	Feldblum T.V., Smith H.O., Venter J.C., Nealson K.H., Fraser C.M.,				
RT	"Genome sequence of the dissimilatory metal ion-reducing bacterium				
RT	Shewanella oneidensis."				
RL	Nat. Biotechnol. 20:1118-1123(2002).				
DR	ENML; A015532; AAN53942.1; -.				
DR	TIGR; SO0866; -.				
KW	Complete proteome.				
SQ	SEQUENCE 139 AA; 14811 MW; 41ECLCFA76957920 CRC64;				
Query Match	14.6%; Score 112; DB 16; Length 139;				
Best Local Similarity	28.3%; Pred. No. 0.14; Length 139;				
Matches	32; Conservative 22; Mismatches 45; Indels 14; Gaps 3;				
QY	39 SGFDSTLSIYQGSANAALQSDARKSETTITQSGYNGADVQGGADNSTIETQNGFR 98				
Db	41 SGRDNLIDLVOQGTANQGVFGSGDNS-AVYTAGNDNISLVITQIGTNEVQLLVGAQ 99				
QY	99 NNATYDQVTRVVTHEMAHAGNNALVNOTASSVVMVROVQFGNNATANQY 151				
Db	100 NKASITQI-----GNDNLVLQNLGS-GNFSIQIADGAAISITQY 139				
RESULT 10					
Q89JI3	Q89JI3	PRELIMINARY;	PRT;	171 AA.	
ID	Q89JI3	PRELIMINARY;	PRT;	171 AA.	
AC	Q89JI3	PRELIMINARY;	PRT;	171 AA.	
DT	01-JUN-2003 (Tremblrel. 24, Created)				
DT	01-JUN-2003 (Tremblrel. 24, Last sequence update)				
DT	01-JUN-2003 (Tremblrel. 24, Last annotation update)				
DE	CsgA protein.				
GN	CSGA OR BLI5300				
OS	Bradyrhizobium japonicum.				
OC	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;				
OC	Bradyrhizobiaceae; Bradyrhizobium.				
OC	NCBI_TaxID=375;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=USDA 110;				
RX	MEDLINE=22484998; PubMed=12597275;				
RA	Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,				
RA	Sasamoto S., Watanabe A., Idesawa K., Iriiguchi M., Kawashima K.,				
RA	Kohara M., Matsumoto M., Shimpo S., Tsuruoka H., Wada T., Yamada M.,				
RA	Tabata S.;				
RT	"Complete genomic sequence of nitrogen-fixing symbiotic bacterium				
RT	Bradyrhizobium japonicum USDA110."				
RL	DNA Res. 9:189-197(2002).				
DR	EMBL; AF005954; BAC50565.1; -.				
KW	Complete proteome.				
SQ	SEQUENCE 171 AA; 17448 MW; 995DB08C01498381 CRC64;				
Query Match	13.7%; Score 105; DB 16; Length 130;				
Best Local Similarity	24.7%; Pred. No. 0.46; Length 130;				
Matches	37; Conservative 27; Mismatches 50; Indels 26; Gaps 4;				
QY	4 LKVAAPAAIVVSGSALAGVVPQWGGGNHNGGNSGPDSTLSIYQVGSANAALQSDA 63				
Db	1 MRITLVATAIALSALTVDQAQ-----AGNSA-----SVLQFGTNSFSIQTS 45				
QY	64 RKSETTITQSGYNGADVQGGADNSTIETL---QNGFRNNATYDQVTRVVTHEMAHAGG 120				
Db	46 TSNATTLQFGATWTATTLTGSLTLVNTAVTCGGTTATASNTALTGGV-----GG 97				
QY	121 NNAALVNTQASDSSVMVROVQFGNNATANQ 150				
Db	98 SNSSLIGQIGANTAGVQGLGILNGSTILQ 127				
RESULT 12					
Q7UCZ1	Q7UCZ1	PRELIMINARY;	PRT;	151 AA.	
ID	Q7UCZ1	PRELIMINARY;	PRT;	151 AA.	
AC	Q7UCZ1	PRELIMINARY;	PRT;	151 AA.	
DT	01-OCT-2003 (Tremblrel. 25, Created)				
DT	01-OCT-2003 (Tremblrel. 25, Last sequence update)				
DT	01-OCT-2003 (Tremblrel. 25, Last annotation update)				
DE	Minor curlin subunit.				
GN	CSGB OR S1108.				
OS					


```

OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN=2457T / ATCC 700930 / Serotype 2a;
RX MEDLINE=22590274; PubMed=12704152;
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
FA Fournier G., Maynew G.F., Plunkett G. III, Rose D.J., Darling A.,
RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
RA Schwartz D.C., Blattner F.R.;
RT "Complete genome sequence and comparative genomics of Shigella
RL flexneri serotype 2a strain 2457T.";
RL Infect. Immun. 71:2775-2786(2003).
DR EMBL: AE016981; AAP16542.1; -.
SQ SEQUENCE 151 AA; 15868 MW; 5D5266B9E64014A0 CRC64;

Query Match 13.6%; Score 104.5; DB 16; Length 151;
Best Local Similarity 29.1%; Pred. No. 0.6;
Matches 34; Conservative 15; Mismatches 57; Indels 11; Gaps 3;

QY 38 SSQPDSLTSLSYQGSNAALALOSDARKSETTITQSGYNGADVGQAGDNSTIELTQNGF 97
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd 21 ARGVDLANSEYNF----AVNELSKSFNQAIIGAQGTNNSAQLRQGSKLLAVVAQEGS 76
QY 98 RNNATYDQLVTRVVVTHEMAH--AGGNNAAALVNQTASDSVMVRQVGFGNNTANQY 151
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd 77 SNRAKIDQ---TG DYNLAYIDQAGSANDASISQGYGNTAMIIQKSGNKANTIQY 129

RESULT 13
QC8CW64 PRELIMINARY; PRT; 160 AA.
ID AC Q8CW64
AC Q8CW64
DT 01-MAR-2003 (TrEMBLrel_23, Created)
DT 01-MAR-2003 (TrEMBLrel_23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel_23, Last annotation update)
DE Minor curlin subunit precursor.
GN CSGB OR C1305.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Maynew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RL of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
DR EMBL: AE016759; AAN79778.1; -.
KW Complete proteome.
SQ SEQUENCE 160 AA; 15963 MW; 49F68448D979B986 CRC64;

Query Match 13.6%; Score 104.5; DB 16; Length 160;
Best Local Similarity 29.1%; Pred. No. 0.65;
Matches 34; Conservative 15; Mismatches 57; Indels 11; Gaps 3;

QY 38 SSQPDSLTSLSYQGSNAALALOSDARKSETTITQSGYNGADVGQAGDNSTIELTQNGF 97
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd 30 ARGVDLANSEYNF----AVNELSKSFNQAIIGAQGTNNSAQLRQGSKLLTVVAQEGS 85
QY 98 RNNATYDQLVTRVVVTHEMAH--AGGNNAAALVNQTASDSVMVRQVGFGNNTANQY 151
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd 86 SNRAKIDQ---TG DYNLAYIDQAGSANDASISQGYGNTAMIIQKSGNKANTIQY 138

RESULT 14
QC83RU7 PRELIMINARY; PRT; 160 AA.
ID AC Q83RU7
```

us-09-543-407-30.rspt

Aug 3 10:54:47 2004

```

Query Match      13.5%; Score 104; DB 16; Length 157;
Best Local Similarity 26.4%; Pred. No. 0.69;
Matches 37; Conservative 20; Mismatches 57; Indels 26; Gaps 4;

QY 12 IVVSGSALAGVVPQWGGGNNHNGGSSGPDSTLSIYQYGSANAALALQSDARKSETTIT 71
Db  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
43 LLPAGAGQAIVIEQQGLG--NEAALDQNGQALLGRIVQSGGAQAYILQ-EGSDLMAIIS 99
QY 72 QSGYCGNGADVGGADNSTIELTQNGFRNNATYDOLVTRVVTHEMAHAGGNNAAALVNQTAS 131
Db  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
100 QQGNNGNSASI-----RQSGSNNAAEQI-----GNDNSASIVQSGS 136

QY 132 DSSVMVRQVGFNNATANQY 151
Db  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
137 GLNSSVTQAGNGQHVQITQY 156

```

Search completed: August 2, 2004, 14:54:42
Job time : 29.7 secs